

UNIVERSIDADE FEDERAL DE SANTA CATARINA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

TESE DE DOUTORADO

UM MODELO PARA MEDIR A QUALIDADE DA INFORMAÇÃO DE *SITES*
UTILIZANDO PROGRAMAÇÃO DIFUSA

Manuel Rosa de Oliveira Lino

ORIENTADOR

Prof. Dr. João Bosco da Mota Alves

Florianópolis

2006

Manuel Rosa de Oliveira Lino

**UM MODELO PARA MEDIR A QUALIDADE DA INFORMAÇÃO DE *SITES*
UTILIZANDO PROGRAMAÇÃO DIFUSA**

**Tese de Doutorado apresentada ao Programa
de Pós-Graduação em Engenharia de Produção
da Universidade Federal de Santa Catarina
como requisito parcial para a obtenção do grau
de Doutor em Engenharia de Produção.**

Orientador: Prof. Dr. João Bosco da Mota Alves

Florianópolis

2006

Lino, Manuel Rosa de Oliveira.

Um modelo para medir a qualidade da informação de *sites* utilizando programação difusa / Manuel Rosa de Oliveira Lino; orientador João Bosco da Mota Alves; co-orientador Rogério Cid Bastos - Florianópolis, 2006.

115 f.

Tese (Doutorado) – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia de Produção, 2006.

Inclui bibliografia

1. Qualidade da informação. 2. Dimensões da qualidade. 3. Regressão logística. 4. Programação difusa. 5. Estatística. I. Alves, João Bosco da Mota. II. Bastos, Rogério Cid. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Produção. IV. Título.

Manuel Rosa de Oliveira Lino

**UM MODELO PARA MEDIR A QUALIDADE
DA INFORMAÇÃO DE SITES
UTILIZANDO PROGRAMAÇÃO DIFUSA**

**Esta tese foi julgada e aprovada para a obtenção do grau
de Doutor em Engenharia de Produção no Programa de
Pós-Graduação em Engenharia de Produção
da Universidade Federal de Santa Catarina.**

Florianópolis, 29 de junho de 2006.

**Prof. Dr. Edson Pacheco Paladini
Coordenador do Programa**

Banca Examinadora

**Prof. Dr. João Bosco da Mota Alves
Universidade Federal de Santa Catarina
Orientador**

**Prof. Dra. Silvia Modesto Nassar
Universidade Federal de Santa Catarina**

**Prof. Dr. Fernando Álvaro O. Gauthier
Universidade Federal de Santa Catarina
Moderador**

**Prof. Dr. Rogério Cid Bastos
Universidade Federal de Santa Catarina
Co-orientador**

**Prof. Dr. Jovane Medina Azevedo
Universidade do Estado de Santa Catarina
Escola Superior de Administração e Gerência**

**Prof. Dr. Jolmar Luis Hawerroth
Universidade Estácio de Sá – Santa Catarina**

Os Versos de Ouro de Pitágoras

Transcritos por seu discípulo Lísias de Tarento séc. VI a.C.

Antes de tudo, presta aos deuses imortais o culto prescrito pela lei. Respeita também a fé jurada. Reverencia em seguida, como convêm, os heróis sublimes e os espíritos semideuses.

Professa o culto da família; cumpre inteiramente deveres para com teu Pai, tua Mãe e todos os teus parentes.

Escolhe para amigo o homem melhor e mais virtuoso. Obedece aos seus doces conselhos e segue o seu exemplo salutar. Esforça-te, tanto quanto possas por não te desviares dele por qualquer pequena falta, porquanto a Vontade marcha lado a lado do Destino como potência diretriz da nossa evolução.

Não te esqueças que deves aprender a dominar as paixões, a ser sóbrio, ativo e casto. Nunca te deixes arrebatado pela cólera.

Sê irrepreensível perante os outros e perante ti próprio. E, acima de tudo, respeita-te a ti mesmo. Que toda a tua vida, que todas as tuas palavras se inspirem na mais pura justiça.

Não te acostumes a viver maquinalmente, mas reflete bem que a morte é o nosso destino comum e que as riquezas materiais podem adquirir-se ou perder-se com a mesma facilidade.

Contra o destino que te haja sido marcado pelas leis divinas, por mais rude que seja não te revoltas contra ele, suporta-o com serenidade, esforçando-te por melhorá-lo quanto possas. Os deuses, com efeito, preservam os virtuosos dos maiores males.

A verdade e o erro encontram-se misturados nas opiniões humanas. Abstem-te, pois, de as aprovar ou rejeitar totalmente, a fim de conservares a tua harmonia. Se o erro triunfa momentaneamente, afasta-te e aguarda com paciência.

Observa sempre com todo o cuidado o que vou dizer-te: não te deixes arrastar sem reflexão pelas palavras e atos de outrem. Fala e procede somente quando a tua razão te houver indicado o mais justo caminho. A deliberação, obrigatória antes da ação, evitar-te-á assim os atos impensados. O que verdadeiramente infelicitava o homem é o falar e proceder sem regra nem medida.

Prevê bem as conseqüências, mesmo as mais afastadas, de cada uma das tuas decisões, para não teres depois de que te arrepender.

Não tenhas a pretensão de fazer o que em realidade ignoras; aproveita, pelo contrário, todas as ocasiões de te instruíres. Assim levarás vida altamente agradável.

É preciso igualmente cuidar da saúde do corpo. Usa moderadamente dos alimentos, das bebidas e dos exercícios que te forem necessários. A justa medida será aquela que impedir que te enfraqueças. Assim, deverás habituar-te a um regime puro e severo.

Segue o teu ideal sem ostentação para não atraíres a incompreensão hostil dos ignorantes.

Não procedas à semelhança das pessoas sem senso que gastam mais do que o necessário ou então se entregam à avareza, mas aprende a guardar em tudo o meio-termo. Não faças, pois, nada que possa prejudicar-te, e por isso reflete antes de proceder.

Logo que acordes, aproveita a lucidez que o sono faculta, para elevar o espírito e refletir nas boas obras que deverás realizar. Todas as noites, antes de adormeceres, faz o exame de consciência. Repassa muitas vezes pelo espírito os atos do dia e pergunta a ti mesmo: o que fiz hoje? Cumpri bem o dever em todas as coisas? Examina assim sucessivamente cada uma das ações. Se compreenderes que andastes mal, repreende-te severamente; se foste irrepreensível, dá-te por satisfeito.

Medita estes conselhos. Ama-os do fundo da alma e esforça-te por pô-los em prática. Conduzir-te-ão às virtudes divinas. Juro-O por Aquele que traçou no nosso espírito a tétrada sagrada, fonte e emblema da Natureza eterna.

Mas, metendo mãos à obra, ora sem cessar aos deuses, para que te ajudem a concluí-la.

Quando estiveres bem compenetrado destes preceitos, chegarás a conceber a constituição íntima dos deuses, dos homens e de todas as coisas, e a compreender a rigor a unidade que penetra toda a obra natural. Conhecerás então esta lei universal: por toda parte, no mundo, a matéria e o espírito são idênticos em natureza.

Tornado assim clarividente, deixarás de ser atormentado por desejos ilegítimos. Reconhecerás então que os homens são os criadores dos seus males. Desgraçados! Não sabem que os seus verdadeiros bens estão ao seu alcance dentro deles mesmos. Como são raros aqueles que conhecem a maneira de se livrar dos tormentos! A cegueira dos homens é tal que lhes perturba a inteligência! Semelhantes a cilindros que rodam ao acaso, não cessam de ser acabrunhados por males infinitos. Porque, não suspeitando a funesta incompreensão que está dentro deles e os acompanha por toda parte, não sabem distinguir o que é justo daquilo que é preciso evitar inteiramente.

Deus, nosso Pai! Digna-te livrá-los dos sofrimentos e mostrar-lhes de que força sobrenatural eles podem dispor! Mas não: aquietemo-los; os homens são a raça dos deuses, e a eles pertence o descobrir as verdades sagradas que a natureza oferece à investigação.

Se chegares a penetrá-las, cumprirás então facilmente todas as minhas prescrições e terás merecido ver-te livre das tuas provações. Mas abstém-te dos alimentos interditos nas purificações e prossegue na obra de libertação da alma, fazendo escolha judiciosa e refletida em todas as coisas, de maneira que estabeleças o triunfo do que há de melhor em ti – o Espírito.

Então, quando abandonares o corpo mortal, elevar-te-ás no éter, deixando de ser mortal, revestirás tu mesmo a forma de um deus imortal.

A G R A D E C I M E N T O S

Ao Prof. João Bosco da Mota Alves, orientador, incentivador, colega e amigo, que tanto me cobrou este trabalho e que desde o começo acreditou que fosse possível realizá-lo.

Ao Prof. Rogério Cid Bastos, co-orientador, colega, amigo e irmão, pelo acompanhamento, críticas e sugestões, mais sugestões, discussões e confrontos, sempre buscando uma melhor resposta aos questionamentos feitos e que, diga-se de passagem, não foram poucos.

À Profa. Silvia Modesto Nassar, colega e amiga, uma das pessoas a quem devo a conclusão deste trabalho, não só por ter assumido parte da carga didática do departamento, mas pela colaboração nas críticas, sugestões e correções ao mesmo que possibilitaram sua conclusão.

Ao Prof. Fernando Álvaro Ostuni Gauthier, colega e amigo, que quando consultado ofereceu suas idéias e opiniões e com seus *insights*.

Ao Prof. Jovane Medina Azevedo, membro da banca, que com suas críticas e sugestões permitiu o enriquecimento do trabalho.

Ao Prof. Jolmar Luis Hawerorth, membro da banca, que soube nas críticas mostrar o caminho para a melhoria do trabalho.

Aos professores Willy Arno Sommer e Ivan Ludgero Ivanqui, pelo apoio e incentivo.

Aos professores dos programas de pós-graduação em Engenharia de Produção e Ciência da Computação, com os quais fiz as disciplinas que me permitiram cumprir os créditos necessários ao programa.

Aos funcionários da secretaria do INE, Ronaldo, Mário, Rita e Carlos pela ajuda, sempre prestimosa.

Aos funcionários da secretaria de pós-graduação do PPGE, Neiva, Rosimeri, Servilho, Sidnei e Juliana, pela colaboração e paciência.

Ao Juliano da Silva de Souza pelo apoio na aplicação dos formulários, ao Henrique Cesar Lemos Jucá pelas revisões e traduções e à Juliana Sell do Vale Pereira pelo trabalho de correção e edição do trabalho final.

Aos professores, aos técnico-administrativos e aos alunos anônimos personagens deste trabalho, que colaboraram com suas respostas, ao formulário de coleta de dados utilizado.

E, finalmente, para não incorrer no erro de esquecer alguém, agradeço a todos que direta ou indiretamente colaboraram neste trabalho.

DEDICATÓRIA E AGRADECIMENTOS

Dedico este trabalho, o mais importante de minha carreira de trinta e três anos de magistério, a todos os colegas, dirigentes ou proprietários das instituições pelas quais passei neste período e que me permitiram e ajudaram a chegar a esta etapa:

Aquela que me deu a formação:

Faculdade de Filosofia Ciências e Letras de Tupã - SP

Aquelas nas quais trabalhei:

Faculdade de Educação de Marília - SP

Faculdade de Ciências Contábeis e de Administração de Tupã - SP

Faculdade de Ciências Contábeis e Administrativas de Rolândia - PR

Faculdade de Filosofia Ciências e Letras de Apucarana - PR

Fundação Faculdade Estadual de Ciências Econômicas de Apucarana - PR

Universidade Estadual de Londrina - PR

Faculdade de Filosofia Ciências e Letras de Dracena - SP

Faculdade de Filosofia Ciências e Letras de Tupã - SP

Universidade Federal de Santa Catarina - SC

Em especial, meus agradecimentos à **Universidade Federal de Santa Catarina**, aos seus dirigentes, servidores professores e técnico-administrativos, alunos e comunidade em geral, com quem convivo há praticamente 25 anos, pela oportunidade de desenvolver minhas atividades num ambiente de pluralidade, respeito e amizade.

Na UFSC, fiz meu mestrado e meu doutorado; minha filha Patrícia aqui se formou em Engenharia de Produção Elétrica; o Manoel Henrique, em Administração e o Maurício está terminando Agronomia.

Obrigado, UFSC.

Este trabalho é dedicado de forma muito especial à minha esposa Silvia, meus filhos Patrícia, Henrique e Maurício. Às Marias, Alice e Salles, e aos saudosos Francisco e Cid.

LISTA DE QUADROS

Quadro 1: Categorias e dimensões do construto de avaliação da qualidade – Wang, Strong.	32
Quadro 2: Categorias e sub características do construto de avaliação da qualidade – Zeist, Hendriks.	33
Quadro 3: Critérios e descrição do construto de avaliação da qualidade – Alexander e Tate.	34
Quadro 4: Categorias e dimensões do construto de avaliação de qualidade – Katerattnakul e Siau.	34
Quadro 5: Nível semiótico, objetivos e dimensões do construto de avaliação da qualidade – Shanks e Corbit.	35
Quadro 6: Categorias e dimensões do construto de avaliação da qualidade – Dedeke.	37
Quadro 7: Tipos de classes de valorização e critérios de qualidade na construção dos construtos – Naum e Rolker.	38
Quadro 8: Tipos de qualidade, categorias e dimensoes do construto de avaliação da qualidade – Kahn, Strong e Wang.	39
Quadro 9: Tipos de qualidade, categorias e dimensões do construto de avaliação da qualidade – Eppler e Muenzenmayer.	40
Quadro 10: Tipos de classe de valorização e critérios de qualidade na construção dos construtos – Moraes e Werneck.	41
Quadro 11: Tipos de dimensão, freqüências e definição da dimensão – Knight e Burn.	43
Quadro 12: Tipos de relações de pertinência para variáveis contínuas – Pacheco	46
Quadro 13: Funções de ligação – Distribuição logística.	56

LISTA DE TABELAS

Tabela 1: Distribuição dos usuários por categoria e sexo.	80
Tabela 2: Descrição da idade por categoria e sexo do usuário.	81
Tabela 3: Estatísticas da variável idade por categoria.	85
Tabela 4: Distribuição dos usuários por local de trabalho-estudo.	86
Tabela 5: Tipo de acesso ao site na UFSC.	86
Tabela 6: Tipo de acesso ao site na Residência.	87
Tabela 7: Frequência de acesso na UFSC por categoria.	87
Tabela 8: Estatísticas das avaliações das dimensões.	88
Tabela 9: Estatísticas da importância dada às dimensões.	89
Tabela 10: Avaliação da percepção da qualidade.	90
Tabela 11: Testes globais para os coeficientes no modelo.	91
Tabela 12: Estimativas dos coeficientes.	91
Tabela 13: Sumário do modelo.	92
Tabela 14: Classificação da variável dependente.	92
Tabela 15: Estimativas dos coeficientes.	93
Tabela 16: Estimativas dos coeficientes do Mod2A – estudantes.	94
Tabela 17: Matriz de correlações dos coeficientes do Mod2A.	94
Tabela 18: Estimativas dos coeficientes do Mod2B – servidores professores.	95
Tabela 19: Matriz de correlações dos coeficientes do Mod2B.	95
Tabela 20: Estimativas dos coeficientes do Mod2C – servidores técnico- administrativos.	96
Tabela 21: Matriz de correlações dos coeficientes do Mod2C.	97

LISTA DOS GRAFICOS E FIGURAS

Gráfico 1: Gráfico da função de pertinência.	51
Gráfico 2: Distribuição dos usuários por categoria e sexo.	81
Gráfico 3: Distribuição das idades pelas categorias.	82
Gráfico 4: Distribuição das idades dos alunos.	83
Gráfico 5: Distribuição das idades dos professores.	83
Gráfico 6: Distribuição das idades dos técnico-administrativos.	84

Figura 1: Mecanismo para recuperar informação sobre a qualidade do site.	67
---	----

RESUMO

Sites inseridos na internet disponibilizam informações científicas, culturais, esportivas, notícias, entretenimento e outras de utilização cada vez mais freqüente no dia-a-dia das pessoas. Os *sites* nem sempre possuem uma qualidade da informação compatível com a tecnologia presente e disponível. O presente estudo tem como objetivo principal apresentar um modelo para avaliar a Qualidade de Informação em *sites*, com base nas dimensões clássicas da qualidade. Agrega as dimensões em grupos e utiliza-se da aplicação de modelos de regressão logit para elaborar um modelo geral de otimização que permita, a partir da definição de conjuntos de dimensões, propor uma solução para maximizar a percepção da qualidade observada em *sites*, pelos usuários. O modelo possibilitará ao decisor investir no grupo de dimensões que proporcionaria um melhor resultado no acréscimo da percepção da qualidade declarada pelo usuário.

Palavras-chave: Qualidade da informação, Dimensões da qualidade, Regressão logit, Programação difusa, Estatística.

A B S T R A C T

Internet websites publishes scientific, cultural, sports, entertainment information, among several other types of information, used more and more frequently every day by the community. These websites usually don't have an information quality that is compatible with the technology available nowadays. This study aims at presenting a model to evaluate Information Quality in websites, based on classical quality dimensions. These dimensions were clustered in groups and used in logit regression models to elaborate a general optimization model to propose a solution that maximizes the perception of quality perceived by users on websites, from the definition of a group of dimensions. This model will allow the policy-maker to invest in the group of dimensions that brings better results towards increasing quality perceived by the user.

Keywords: Information Quality, Quality Dimensions, Logit Regression, Fuzzy Programming, Statistics

SUMÁRIO

1 INTRODUÇÃO	18
1.1 ORIGEM DO TRABALHO	18
1.2 PROBLEMA DE PESQUISA.....	19
1.3 OBJETIVOS DO TRABALHO.....	20
1.3.1 Objetivo geral.....	20
1.3.2 Objetivos específicos.....	20
1.4 MÉTODOS.....	21
1.5 DELIMITAÇÃO DO TRABALHO	22
1.6 ESTRUTURA DO TRABALHO	23
2 QUALIDADE DA INFORMAÇÃO	25
2.1 DEFINIÇÃO DE QI	28
2.2 FERRAMENTAS PARA AVALIAR QI.....	28
2.3 QI COMO UMA SÉRIE DE DIMENSÕES.....	30
2.3.1 Ferramenta conceitual para QI - Wang e Strong.....	31
2.3.2 Modelo de Zeist e Hendriks.....	32
2.3.3 QI para ambientes <i>web</i> – Alexander e Tate.....	33
2.3.4 QI para <i>sites</i> – Katerattnakul e Siau	34
2.3.5 Ferramenta para QI baseada na semiótica - Shanks e Corbit	35
2.3.6 Ferramenta conceitual para medir qualidade de sistemas de informação - Dedeker	35
2.3.7 QI a partir de meta critérios - Naumann e Rolker.....	37
2.3.8 Mapeamento de QI em um modelo PSP/IQ - Kahn <i>et al.</i>	38
2.3.9 Ferramentas conceituais para QI no contexto de <i>WebSites</i> - Eppler e Muenzenmayer	39
2.3.10 Uma abordagem de AQ de aplicações <i>web</i> - Moraes e Werneck	40
2.4 QI EM SITES E O PROCESSO COGNITIVO.....	42
3 CONJUNTOS DIFUSOS.....	44
3.1 INTRODUÇÃO.....	44
3.2 CARACTERÍSTICAS GERAIS DOS CONJUNTOS DIFUSOS	45

3.3 OPERADORES PADRÃO	46
3.3.1 União	47
3.3.2 Intersecção	48
3.3.3 Complementação	49
3.4 PROBLEMAS DE OTIMIZAÇÃO DIFUSA	49
3.5 PROBLEMAS DE OTIMIZAÇÃO COM FUNÇÕES E RESTRIÇÕES DIFUSAS.....	50
 4 A REGRESSÃO	 53
4.1 INTRODUÇÃO.....	53
4.2 REGRESSÃO LOGÍSTICA - LOGIT	53
4.3 O MODELO DE REGRESSÃO LOGÍSTICA	54
4.4 CRITÉRIOS PARA AVALIAR AJUSTE DE UM MODELO	56
4.5 INTERPRETAÇÃO DOS COEFICIENTES	57
4.6 REGRESSÃO LOGÍSTICA E O PACOTE SPSS	58
 5 O MODELO PROPOSTO.....	 64
5.1 INTRODUÇÃO.....	64
5.2 OBTENDO DIMENSÕES DA QUALIDADE RELEVANTES EM UM SITE	65
5.3 DESENVOLVENDO MECANISMOS PARA PRIORIZAR DIMENSÕES	69
5.4 IDENTIFICAÇÃO DAS CARACTERÍSTICAS E DIMENSÕES RELEVANTES POR MODELOS DE REGRESSÃO	70
5.4.1 Mod1 – Considerando as categorias dos usuários.....	71
5.4.2 Mod2 - Considerando os grupos de dimensões	71
5.4.3 Mod3 - Considerando os grupos de dimensões e as categorias	72
5.5 DEFINIÇÃO DOS MODELOS DE OTIMIZAÇÃO POR PROGRAMAÇÃO DIFUSA	72
5.5.1 Otimização do processo de avaliação da qualidade	72
5.5.2 Construção dos modelos de otimização difusa	73
5.6 ANÁLISE DOS RESULTADOS	75
 6 APLICAÇÃO DO MODELO	 77
6.1 CONSTRUÇÃO DO INSTRUMENTO DE COLETA DOS DADOS.....	77
6.2 APLICAÇÃO DO INSTRUMENTO	78
6.3 RESULTADO PRELIMINAR	79

6.3.1 Quanto às características dos usuários.....	80
6.3.2 Quanto à avaliação das dimensões	88
6.3.3 Quanto à importância dada às dimensões	89
6.3.4 Quanto à percepção da qualidade	90
6.4 APLICAÇÃO DOS MODELOS.....	90
6.4.1 Mod1 – Considerando a categoria e o sexo do cliente	90
6.4.2 Mod2 - Considerando os grupos de dimensões	91
6.4.3 Mod3 - Considerando os grupos de dimensões e as categorias	93
6.5 MODELO DE OTIMIZAÇÃO DIFUSA	97
6.6 DISCUSSÃO DOS RESULTADOS	99
 7 CONCLUSÕES E RECOMENDAÇÕES	 100
7.1 CONCLUSÕES.....	100
7.2 RECOMENDAÇÕES PARA FUTUROS TRABALHOS	103
 8. REFERÊNCIAS	 104
Apêndice	111

1 INTRODUÇÃO

O surgimento da *internet* em nível mundial pode ser considerado como tendo ocorrido no final da década de 60, tendo a primeira conexão internacional da ARPANET¹ acontecido entre a *University College of London* (Inglaterra) e o *Royal Radar Establishment* (Noruega) em 1973².

A *internet*, mais do que bem-sucedida, provocou uma ruptura construtiva com o passado, sendo uma tecnologia integradora para onde todas as formas de comunicação e entretenimento poderão rapidamente migrar.

Uma grande preocupação a ser discutida é a necessidade de se garantir a acessibilidade, a usabilidade e a qualidade das informações dos *sites* na *internet*. É importante dar condições de navegação e de uso dos recursos do *site* a qualquer pessoa, a fim de garantir a universalização dos serviços e a inclusão digital.

1.1 ORIGEM DO TRABALHO

Definir Qualidade da Informação (QI) implica um processo complexo e multifacetado, que se tornou ainda mais difícil no contexto de busca de informações de fontes não validadas, como aquelas que compõem os *sites* na *World Wide Web*. Neste trabalho propõe-se discutir um modelo para avaliar QI em *sites* tendo como base as dimensões mais clássicas, as quais

¹ http://server.berkeley.edu/virtual-berkeley/email_history

² <http://www.ocean.ic.net/ftp/doc/nethist.html>

podem ser aplicadas para mensurar o conceito de QI no contexto do seu uso. Todavia, a compreensão de QI do ponto de vista do usuário também implica a compreensão dos processos envolvidos na percepção da informação na *web* antes da aplicação das métricas para a avaliação da qualidade.

Assim, uma abordagem para a mensuração da qualidade da percepção é proposta. Ela engloba características do usuário, do ambiente e da tarefa; quantificação das dimensões de qualidade dentro do contexto de usuário, ambiente e tarefa; implementação de um processo para avaliar a qualidade e de um mecanismo de *feedback* para continuamente refinar o aperfeiçoamento do processo de busca de qualidade baseado na relevância.

Pretende-se, ao mesmo tempo, buscar uma maneira de refinar ainda mais os critérios de QI e o desenvolvimento de interfaces para o usuário de maneira que estas possam mensurar a aceitação e a satisfação do mesmo com o processo de busca de qualidade da informação no *site*.

1.2 PROBLEMA DE PESQUISA

O crescimento da *internet*³ como um ambiente para troca de informações e a ausência de padrões rígidos quanto ao conteúdo destas informações conduzem a inúmeros problemas de qualidade da informação.

Uma questão central é a incapacidade de as tecnologias de busca (máquinas de busca) vagarem por uma imensidão de informações e retornarem uma informação de qualidade para uma simples consulta de um usuário. Contudo, a carência de padrões estabelecidos tem resultado em

³ <http://www.mit.edu/people/mkgray/net/web-growth-summary.html>

freqüentes problemas de qualidade de informação (EPPLER, MUENZENMAYER, 2002).

Neste trabalho, são tratadas questões envolvidas na determinação do que é qualidade e como este conceito está associado à obtenção da informação na *internet* com qualidade para o usuário final. O foco do trabalho está nos conceitos de qualidade da informação.

1.3 OBJETIVOS DO TRABALHO

1.3.1 Objetivo geral

O objetivo geral deste é o de propor um modelo para avaliar a percepção da qualidade de um *site* com base em características individuais e dimensões pelas quais a informação pode ser interpretada.

1.3.2 Objetivos específicos

São objetivos específicos:

- a. Identificar formas de avaliar como características individuais podem influenciar na percepção da qualidade do *site*;
- b. Identificar dimensões de qualidade que devem ser consideradas na análise da QI de um *site*;
- c. Identificar como as dimensões de qualidade influenciam na qualidade global de um *site*;
- d. Estabelecer níveis críticos que permitam a adoção de políticas visando maximizar a QI, sujeita a restrições técnicas e orçamentárias;
- e. Desenvolver um estudo piloto para validar o modelo proposto.

1.4 MÉTODOS

O trabalho contempla o desenvolvimento de um modelo no qual são identificadas as características de qualidade da informação que agregam valor a um *site*. Em paralelo, é avaliado como características individuais podem influenciar na percepção desta qualidade.

São freqüentes os modelos para medir a influência de características pessoais e técnicas que ampliam a percepção da qualidade em um *site*. Estas influências são maximizadas utilizando-se conceitos oriundos de conjuntos difusos.

Este contém, ainda, uma aplicação prática. Trata-se de uma pesquisa de caráter descritivo-exploratório realizada mediante um estudo de caso. O caso analisado foi o *site* da Universidade Federal de Santa Catarina, www.ufsc.br.

Será desenvolvido um instrumento de coleta de dados que permitira avaliar a percepção de qualidade dos usuários do *site*.

A aplicação seguirá os procedimentos estatísticos de amostragem considerando três sub-populações características, Alunos, Servidores Professores e Servidores Técnico-administrativos, STA.

A fórmula a ser utilizada para a definição do tamanho da amostra:

$$n = \frac{p * (1 - p) * z^2}{e^2} \quad \text{Equação 1}$$

onde:

p : é a proporção amostral de usuários que afirma a qualidade do *site* como boa; p é o estimador de π (que representa a proporção de usuários que afirmam qualidade como boa na população);

q : é a proporção amostral de usuários que afirma não à qualidade do *site* ou seja qualidade ruim;

z o valor obtido a partir da curva normal para uma dada confiança;

e o valor máximo do erro amostral permitido para a estimação de π ;

n o tamanho da amostra;

A partir das informações levantadas pelo instrumento de coleta os dados serão tratados e uma caracterização da amostra será apresentada.

Os grupos de dimensões da qualidade identificados serão processados através de modelos de regressão logit (ou logística) e a análise no modelo final ocorrerá por otimização difusa.

1.5 DELIMITAÇÃO DO TRABALHO

Este trabalho está centrado na percepção da qualidade que um usuário apresenta em relação a um *site*. Não se discutem formas de elaboração do *site*, de modo a previamente criar conclusões básicas para QI. Questões associadas a usabilidade de *sites* e sua implicação decorrentes de modificações em dimensões de qualidade da informação, sob o ponto de vista técnico, também não são consideradas. O enfoque básico do trabalho é o usuário final.

1.6 ESTRUTURA DO TRABALHO

A estrutura desta tese é composta por sete capítulos, os quais foram assim divididos no intuito de proporcionar uma melhor compreensão do assunto.

No primeiro capítulo é feita a introdução, apresentando origem e problema do assunto proposto, além dos objetivos e métodos, incluindo a delimitação do trabalho.

No segundo capítulo, efetua-se a fundamentação teórica com base em autores que já pesquisaram sobre o tema aqui abordado. Desta forma, são apresentados aspectos teóricos sobre Qualidade da Informação em suas diversas abordagens e conceitos. Neste capítulo são apresentadas também as principais ferramentas para avaliar a qualidade da informação.

No terceiro capítulo, é apresentado um estudo sobre conjuntos difusos e as técnicas de programação difusa que serão utilizadas na otimização dos recursos para melhorar o conceito de qualidade de um *site*.

No quarto capítulo, é apresentado um resumo sobre a técnica de regressão logística, são propostos modelos de regressão para verificar a percepção da qualidade em *sites*, e há ainda um resumo do processamento de modelos de regressão através de um programa estatístico, o SPSS - *Statistical Package for the Social Sciences - Release 11.5.0*.

No quinto capítulo apresenta-se, além dos modelos de regressão que possibilitam identificar características e dimensões relevantes da percepção da qualidade, o modelo de otimização definido a partir da análise dos

coeficientes da regressão, modelo que será obtido através da programação difusa.

No sexto capítulo, os modelos de regressão e otimização são processados e os resultados apresentados e comentados com base nas teorias apresentadas nos dois últimos capítulos. Este estudo pode ser aplicado a *sítes* em que se busca um incremento em dimensões ou grupos de dimensões da qualidade percebida pelos usuários quando houver limitação de recursos humanos, financeiros de tempo ou outro qualquer.

Por fim, no sétimo capítulo, de conclusões e recomendações, resgatam-se, além dos principais aspectos descritos durante a realização deste trabalho, as dificuldades encontradas, além de sugestões e recomendações para a realização de eventuais trabalhos paralelos ou complementares.

O encerramento apresenta as referências consultadas para a realização deste trabalho.

2 QUALIDADE DA INFORMAÇÃO

Tomando como principal referencial as categorias analíticas de Gaston de Bachelard - estádios de um conceito e obstáculo epistemológico -, a leitura da noção "qualidade da informação", tal como é abordada na literatura, revela que se trata de uma noção vaga, imprecisa, situando-se muito próxima ao entendimento do senso comum. As definições geralmente baseiam-se na adscrição de atributos passíveis de mensuração. A tendência dominante é a de se imprimir um julgamento de valor positivo à informação, relegando-se a um plano secundário seu lado negativo. A desconstrução das noções relativas à avaliação da informação constitui-se, pois, em um passo necessário para propiciar o redirecionamento da construção conceitual, de modo a sintonizar-se com as exigências do novo momento tecnológico e social. (NEHMY, PAIM, 1998)

Na sociedade da informação, a *internet* assume papel de preponderante importância. A quantidade e o descontrole das informações nela contidas, a troca de informações que nela se processa, moldam comunidades eletrônicas, incrementando a comunicação pessoal e profissional sem fronteiras.

A proliferação de recursos na *internet* possibilita uma informação-desinformação a qual encerra em si não apenas uma comunidade virtualmente comunicável e memorável, mas coletivos humanos heterogêneos. As informações podem ser de procedência ou qualidade duvidosa, dado que poucos dados na rede foram referenciados ou revistos, e a ausência de controle da informação que flui na *web* enfatiza a necessidade de instrumentalizar seus usuários para que o processo incorpore a dimensão de qualidade da informação.

A qualidade da informação pode ser traduzida em informações íntegras, atualizadas e precisas *on-line*, e de preferência *just-in-time*, que permitam uma ideal tomada de decisão. Não se trata apenas de uma questão de força, isto é, de poder de processamento e sofisticação de

software e *hardware*, e sim de uma questão de inteligência, ou seja, da habilidade em transformar a massa de dados operacionais em informações consistentes, que agreguem valor ao negócio.

Agregar valor ao negócio pode ter diferentes leituras ou reunir funções distintas, mas que se complementam. Primeiro, suportar da melhor forma possível o processo de tomada de decisão com informações de qualidade.

A informação agrega valor quando permite à empresa perceber oportunidades e ameaças à sua operação, detectando problemas, ou melhor, oportunidades. No caso de um usuário comum, atende à sua busca de uma maneira ampla, permitindo uma tomada de decisão.

Mais ainda, num mundo competitivo como o atual, informação sobre o ambiente interno e externo, os clientes e o mercado é fundamental para a sobrevivência da organização. Possuir inteligência competitiva é ter o tempo todo o mercado, atual e prospectivo, bem monitorado.

E ainda, talvez na sua função mais nobre, a informação agrega valor ao negócio quando da sua análise chega-se a novas maneiras de fazer negócio, novos serviços e novos produtos, de forma a estabelecer a perenidade do negócio e dar uma qualidade de vida adequada aos funcionários. A informação permite compartilhar conhecimentos entre grupos ou pessoas de trabalhos cujos conteúdos demorariam a ficar disponíveis pelos processos convencionais de publicação.

Não existe praticamente nenhuma profissão que não se tenha beneficiado das informações presentes na *World Wide Web* - *WWW*. Mas também é possível afirmar que muitos usuários tiveram enormes contratempos por informações obtidas através da rede e que os se

mostraram erradas ou manipuladas por indivíduos inescrupulosos com objetivos de auferir rendimentos ou recursos indevidamente.

No caso específico deste trabalho, a revolução nos paradigmas de produção do conhecimento e as implicações na qualidade de informação disponibilizadas na *World Wide Web* são analisadas. Destacam-se aqui ainda os estudos e as iniciativas nacionais e internacionais desenvolvidos com o objetivo de avaliar, controlar e assegurar a qualidade da informação na *web*.

A quantidade de informação na *web* vem aumentando em ritmo surpreendente e os processos de armazenamento, recuperação e disseminação de informação ali encontram um espaço privilegiado para seu acesso e intercâmbio, estando ao alcance dos mais diferentes públicos. Estas informações que instituições públicas, privadas, indústrias, governos, associações e indivíduos agregam diariamente na *internet*, muitas vezes de modo desorganizado e não necessariamente científico, tornam necessária a utilização de critérios de qualidade que diferenciem e legitimem a informação relevante do lixo eletrônico.

O rápido, simultâneo e crescente progresso nas tecnologias de informação apresenta caminhos inovadores e começa a ultrapassar os processos tradicionais de publicação de pesquisa e outras formas convencionais de difusão de conhecimento. Uma abordagem moderna de gestão da qualidade da informação pode fortalecer as redes de pesquisa através da *internet* e de outros meios de comunicação, além de construir novas redes que permitirão o rápido compartilhamento de conhecimento e de experiências práticas entre as mais diversas profissões, organizações e comunidades virtuais no mundo.

2.1 DEFINIÇÃO DE QI

Qualidade da Informação (QI) ou Qualidade de Dados (QD) é comumente associado com um conceito multidimensional (KLEIN, 2001), estando a interpretação das características atribuídas de acordo com o ponto de vista filosófico do autor. De forma mais comum, a expressão “Qualidade de Dados” é interpretada como sendo os dados que estão “ajustados para o uso” (WANG, STRONG, 1996), o que implica uma relatividade, uma vez que o que está adequado para o uso de um pode não estar adequado para o uso de outro (TAYI, BALLOU, 1998).

Qualidade é definida pela norma técnica NBR ISO 8402 de 1994 (Gestão da qualidade e garantia da qualidade) como a totalidade das características de uma entidade que lhe confere a capacidade de satisfazer as necessidades explícitas e implícitas. Entidade é o produto propriamente dito, as necessidades explícitas são as próprias condições e objetivos propostos pelo produtor, e as necessidades implícitas são condições mais subjetivas, como as diferenças entre as necessidades dos usuários, a evolução no tempo, as implicações éticas, as questões de segurança e outras.

Nesse contexto, este trabalho vai procurar estabelecer critérios de análise de avaliação de qualidade para aplicações *web*.

2.2 FERRAMENTAS PARA AVALIAR QI

A baixa qualidade das informações QI distribuídas através de fontes autônomas é um dos maiores problemas para os consumidores de informação. Independentemente do consumidor, desde usuários casuais de

serviços de informação da *World Wide Web*, estudantes em busca de subsídios para seus trabalhos, profissionais das mais diversas áreas do conhecimento ou tomadores de decisão utilizando uma intranet para obter informações de diferentes departamentos, todos estão sujeitos ao grande problema que é a baixa qualidade das informações de alguns *sítes*.

A necessidade de medidas contra a baixa qualidade é clara e muitos projetos têm buscado propor métodos com o intuito de aumentar a qualidade da informação e dos dados. No entanto, muitas abordagens pecam pela falta de métodos ou sugestões de como avaliar os escores de classificação, além do que a avaliação de QI é considerada difícil por motivos como:

1. Os critérios de QI são geralmente de natureza subjetiva e não podem, portanto, ser avaliados automaticamente, por exemplo, independentemente do usuário;
2. As fontes de informação são em sua maioria autônomas e muitas vezes não publicam metadados⁴ de qualidade utilizável, muitas inclusive tomam medidas para evitar a avaliação da qualidade da informação disponibilizada;
3. A massa enorme de dados a serem avaliados impede a avaliação de todo o conjunto de informações, assim, técnicas de amostragem muitas vezes são necessárias, o que diminui a precisão da classificação avaliada;
4. A informação proveniente de fontes autônomas algumas vezes é sujeita a mudanças surpreendentes em conteúdo e qualidade.

⁴ A mais breve definição de metadados é "dado sobre dado". Neste contexto, metadado refere-se a alguma estrutura descritiva da informação sobre outro dado, usado para ajudar na identificação, descrição, localização e gerenciamento de recursos da *web*.

Finalmente, pode-se afirmar que a fonte mais confiável para a classificação da QI deveria ser a própria fonte de informação. Todavia, tais metadados muitas vezes não estão disponíveis, especialmente se a fonte compete com outras. Desta maneira, devem ser desenvolvidas outras metodologias que avaliem independentemente metadados de QI de uma maneira eficiente. A avaliação deve ser tão automatizada quanto possível, mas ainda ser tão orientada pelo usuário quanto necessário.

2.3 QI COMO UMA SÉRIE DE DIMENSÕES

Diversos autores durante a última década de pesquisas em Sistemas de Informação buscaram estabelecer referenciais aceitos em QI. Dentre os quais pode-se citar: Wang e Strong (1996), Zeist e Hendriks (1996), Alexander e Tate (1999), Katerattanakul e Siau (1999), Shanks e Corbitt (1999), Dedeker (2000), Reis, et al., (2002), Olson (2003), Naumann e Rolker (2000), Klein (2002), Gertz et al. (2004), Garrity e Sanders (2006), Devanshu (2002), Cappiello et al., (2004), Bouzeghoub e Peralta (2004), Bacharach (2002).

Enquanto variam na abordagem e na aplicação, os referenciais compartilham uma série de características relacionadas com as dimensões de qualidade. Dentre os múltiplos trabalhos apresentados, serão mostrados e discutidos alguns que sintetizam o preceito geral de qualidade.

A QI pode ser vista sob a ótica da acessibilidade. Entretanto, também pode ser vista sob o aspecto de sua representação, atualidade etc. Assim tem-se que a QI está associada a diversas dimensões. Com base nestas diferentes dimensões, vários modelos para defini-la são encontrados na

literatura. A seguir são descritos os principais estudos realizados neste campo.

2.3.1 Ferramenta conceitual para QI - Wang e Strong

Em seu trabalho, Wang e Strong (1996) propõem uma ferramenta conceitual para qualidade da informação. De acordo com os autores, a baixa qualidade dos dados pode ter impacto substancial na esfera social e econômica. Apesar de as empresas estarem melhorando a qualidade dos dados através de medidas práticas e ferramentas, estes esforços tendem a se concentrar no aspecto da precisão.

Segundo Wang e Strong, os consumidores de informação têm uma conceitualização mais generalizada da qualidade dos dados que os profissionais de Sistemas de Informação. Com base nestes pontos, propõem uma estrutura que captura os aspectos da qualidade da informação importantes aos consumidores. A pesquisa é dividida em estágios, de modo a desenvolver uma estrutura hierárquica para organizar as dimensões da qualidade da informação.

Os autores estruturam a avaliação da qualidade através de quatro categorias e quinze dimensões, conforme apresentado no Quadro 1.

C o n s t r u t o	
Categoria	Dimensões
Intrínseca → Própria da informação que está sendo buscada ou gerada	Precisão, Objetividade, Credibilidade e Reputação
Acessibilidade → Quanto à forma de obtenção da informação	Acessibilidade e Segurança
Contextual → Relacionada à tarefa que se pretende realizar	Relevância, Valor adicionado, Atualidade, Completude e Quantidade de informação
Representacional → Quanto à forma de representar e ver a informação	Interpretabilidade, Facilidade de compreensão, Representação concisa e Representação consistente

Quadro 1 – Categorias e dimensões do construto de avaliação da qualidade.
Fonte: Wang e Strong, 1996

2.3.2 Modelo de Zeist e Hendriks

Zeist e Hendriks (1996) utilizam o modelo ISO estendido e consideram que QI pode ser descrita conforme as características e sub-características apresentadas no Quadro 2.

C o n s t r u t o	
Características	Sub-características
Funcionalidade	Adequabilidade, Precisão, Interoperabilidade, Complacência, Segurança e Rastreabilidade
Confiabilidade	Maturidade, Recuperabilidade; Disponibilidade, Degradabilidade e Tolerância a falhas
Eficiência	Comportamento tempo e Comportamento fonte
Usabilidade	Compreensibilidade, Aprendizagem, Operabilidade, Luxúria, Clareza, Auxiliadora, Explícita, Customizável e Amigável
Manutenção	Analisabilidade, Portabilidade, Estabilidade, Testabilidade, Re-usabilidade e Capacidade de realizar manutenção
Portabilidade	Adaptabilidade, Conformação, Substituibilidade e Instantaneabilidade

Quadro 2 – Características do construto de avaliação da qualidade.
Fonte: Zeist e Hendriks, 1996

2.3.3 QI para ambientes *web* – Alexander e Tate

O trabalho proposto por Alexandre e Tate (1999) introduz ferramentas para avaliar qualidade para ambientes *web*. Nele, os autores utilizam a noção de construtos distribuídos em diversos critérios. O Quadro 3 sumariza os critérios dos construtos utilizados e seus respectivos objetivos.

C o n s t r u t o	
CrITÉRIOS	Explicação
Autoridade	Informação convalidada – autor visível
Precisão	Informações confiáveis – livres de erros
Objetividade	Apresentada sem vieses pessoais
Atualidade	Conteúdo atualizado
Orientação	Claramente identificada a audiência-alvo
Navegabilidade	Design intuitivo

Quadro 3 – Critérios e descrição do construto de avaliação da qualidade.
Fonte: Alexander e Tate, 1999

2.3.4 QI para *sítes* – Katerattnakul e Siau

O trabalho de Katerattnakul e Siau (1999) enfoca especificamente QI em *sítes*. Como Alexandre e Tate, também se utilizam de construtos. Estruturalmente para cada categoria, os autores enumeram diferentes dimensões de QI no *site*. O Quadro 4 apresenta um resumo de seu trabalho.

C o n s t r u t o	
Categoria	Dimensões
Intrínseca →	Precisão e erros de conteúdo, Correção, Trabalhável com hiperlinks relevantes
Contextual →	Fornece informações sobre o autor
Representacional →	Organização, Características visuais, Características tipográficas, Consistência, Atratividade e Intensividade
Acessibilidade →	Fornece ferramentas de navegação

Quadro 4 – Categoria e dimensões do construto de avaliação da qualidade.
Fonte: Katerattnakul e Siau, 1999

2.3.5 Ferramenta para QI baseada na semiótica - Shanks e Corbit

A semiótica é uma teoria baseada nas instruções por sinais. Shanks e Corbit (1999) estabeleceram dimensões de QI, considerando diferentes níveis e objetivos centrados nas dimensões segundo uma abordagem semiótica. O Quadro 5 mostra o construto, classificado por dimensões, objetivos e nível semiótico, proposto pelos autores.

C o n s t r u t o		
Nível Semiótico	Objetivos	Dimensões
Sintático	Consistência	Bem definido e Sintaxe formal
Semântico	Preciso e Completo	Compreensivo, Sem ambigüidades, Significante e Correto
Pragmático	Usável e Útil	Atualizado, Conciso, Facilmente acessível e Reputável
Social	Compartilhando compreensão de significado	Compreensão e Ausência de viés

Quadro 5 – Nível semiótico, objetivos e dimensões do construto de avaliação da qualidade.

Fonte: Shanks e Corbit, 1999

2.3.6 Ferramenta conceitual para medir qualidade de sistemas de informação - Dedeke

Para Dedeke (2000), a mensuração da qualidade dos dados não pode ser desenvolvida isolada dos processos de geração dos dados e dos contextos dentro da qual é utilizada. O autor apresenta categorias de qualidade para Sistemas de Informação (SI). A premissa é que a qualidade de SI não pode, ou ao menos não deveria, ser considerada isoladamente da

qualidade dos dados, das características do trabalho e dos processos envolvidos durante a utilização dos dados.

A disponibilidade e a velocidade de aquisição, processamento e transferência de informação têm se tornado entidades estratégicas nas organizações (PASTERNAK, VISCIO, 1998). A importância estratégica da informação leva à idéia de que as informações deveriam ser gerenciadas tal qual se gerencia um produto (WANG, LEE, PIPINO e STRONG, 1998).

O trabalho mostra que a estrutura para o desenvolvimento de medidas de qualidade para SI deve ser iniciada considerando-se as necessidades para que este sistema será utilizado na produção, no gerenciamento ou processamento de informações e dados.

Dois passos pelo menos são necessários (WANG *et al.*, 1998), para criar tal estrutura. Inicialmente, é necessário definir de forma clara o que qualidade de SI significa para uma organização. A seguir desenvolvem-se as dimensões para as categorias de qualidade identificadas.

A estrutura proposta por Dedeké foi conceitualizada a partir do ponto de vista dos consumidores de dados e limitada em escopo, focada nos dados, trabalhos e eventos ligados a soft/hardware. Não foram incluídas questões culturais (GARRITY e SANDERS, 1998), nem se estende a questões comportamentais (LAU e HORVITZ, 1999). A estrutura pode ser vista no Quadro 6:

C o n s t r u t o	
Categoria	Dimensões
Ergonomia	Facilidade de Navegação, Conforto, Aprendizagem; Sinais visuais e Sinais auditivos
Acessibilidade	Acesso técnico, Disponibilidade dos sistemas, Segurança técnica, Acessibilidade aos dados, Compartilhamento de dados e Convertibilidade de dados
Transacional	Controlabilidade, Tolerância a erros, Adaptabilidade, Sistemas de feedback, Eficiência e Sensibilidade
Contextual	Valor agregado, Relevância, Atualidade, Completude e Dados apropriados
Representação	Interpretabilidade, Consistência, Concisão, Estrutura, Legibilidade e Contraste

Quadro 6 – Categorias e dimensões do construto de avaliação da qualidade.
Fonte: Dedeke, 2000

2.3.7 QI a partir de meta critérios - Naumann e Rolker

A qualidade da informação é um dos aspectos mais importantes da integração de informações na *internet*. Muitos projetos reconhecem isso, e abordam essa realidade através da coleta e classificação de critérios de QI. Dificilmente estes projetos se referem à imensa dificuldade de avaliar a qualificação para estes critérios. Esta tarefa deve preceder qualquer utilização dos critérios para qualificação e integração das informações.

Naumann e Rolker (2000) buscam, como em outros trabalhos de classificação de critérios de QI, a classificação através de uma abordagem orientada na qualificação dos mesmos. Assim, foram identificadas três fontes para metadados de QI. O Quadro 7 mostra as três fontes e os critérios de qualidade do modelo apresentado pelos autores.

C o n s t r u t o	
Classes de Valorização	Crítérios de Qualidade
Sujeito	Credibilidade, Interpretação concisa, Interpretabilidade, Relevância, Reputação, Compreensividade e Valor agregado
Objeto	Completude, Suporte ao cliente, Documentação, Objetividade, Preço, Confiabilidade, Segurança, Atualidade e Verificabilidade
Processo	Precisão, Quantidade de dados, Disponibilidade, Representação consistente, Latência e Tempo de resposta

Quadro 7 – Tipos de classes de valorização e critérios de qualidade na construção dos construtos.

Fonte: Naumann e Rolker, 2000

2.3.8 Mapeamento de QI em um modelo PSP/IQ - Kahn *et al.*

De acordo com Kahn *et al.* (2002), Qualidade de Informação é uma ciência inexata em termos de avaliação e padrões. Apesar de muitos aspectos já terem sido investigados, ainda existe uma necessidade de uma metodologia que avalie quão bem organizações desenvolvem produtos de informação e entregam serviços de informação para os consumidores. Padrões desenvolvidos a partir de tais metodologias podem auxiliar na comparação da qualidade de informação entre organizações, provendo uma linha de base para avaliar melhorias em QI.

A proposta do trabalho de Kahn *et al.* está centrada em um modelo conceitual denominado 2x2 (dois-por-dois), no qual as colunas apreendem a qualidade em conformidade com as especificações como excedente às

expectativas dos consumidores, e as linhas apreendem a qualidade dos produtos e aspectos dos serviços.

O Quadro 8 mostra as dimensões, a classificação e o tipo de qualidade para o mapeamento da qualidade da informação segundo Kahn *et al.*

C o n s t r u t o		
Tipo de Qualidade	Classificação	Dimensões
Produto	Informação perceptível	Livre de erros, Concisa, Representação, Completude e Consistência
	Informação útil	Quantidade apropriada, Relevância, Compreensibilidade, Interpretabilidade e Objetividade
Serviço	Informação segura	Atualidade e Segurança
	Informação utilizável	Credibilidade, Acessibilidade, Facilidade de manipulação, Reputação e Valor agregado

Quadro 8 – Tipos de qualidade, categorias e dimensões do construto de avaliação da qualidade.
Fonte: Kahn, Strong e Wang, 2002

2.3.9 Ferramentas conceituais para QI no contexto de *WebSites* - Eppler e Muenzenmayer

Segundo Eppler e Muenzenmayer (2002), diversas ferramentas poderosas estão disponíveis hoje em dia para avaliar a qualidade da informação no contexto *web*. Estas podem ser categorizadas em cinco tipos diferentes de ferramentas, especificamente:

- sistemas de monitoramento de performance;
- analisadores de *sites*;
- medidores de tráfego;
- ferramentas de mineração *web*;

- ferramentas de pesquisa (que geram informação baseada na opinião do usuário).

Os autores fizeram uso combinado destas ferramentas a fim de permitir a organização para mensurar as múltiplas dimensões da qualidade da informação na *internet* ou no contexto intranet além de demonstrarem quais critérios de qualidade da informação podem ser mensurados com o auxílio destas ferramentas. O Quadro 9 apresenta a classificação adotada pelos autores para tratar qualidade da informação no contexto de *websites*.

C o n s t r u t o		
Tipo de Qualidade	Categorias	Dimensões
Conteúdo	Informação relevante	Compreensibilidade, Precisão, Clareza e Aplicabilidade
	Informação perceptível	Concisa, Consistente, Correta e Atualizada
Mídia	Processo otimizado	Conveniente, Atualidade, Rastreável e Interativa
	Infra estrutura confiável	Acessibilidade, Segurança, Manutenção e Rapidez

Quadro 9 – Tipos de qualidade, categorias e dimensões do construto de avaliação da qualidade.

Fonte: Eppler e Muenzenmayer, 2002

2.3.10 Uma abordagem de AQ de aplicações *web* - Moraes e Werneck

Moraes e Werneck (2000) afirmam que os sistemas de avaliação de aplicativos na *web* devem ser desenvolvidos com enfoque forte no usuário e terem um ciclo de vida rápido e uma grande preocupação com acessibilidade.

O modelo proposto por eles para a análise da qualidade foi subdividido em duas partes: modelo de qualidade para as características externas (percebidas pelos usuários) e internas (percebidas pelos construtores) e modelo de qualidade para qualidade em uso.

O Quadro 10 mostra os requisitos de qualidade definidos para o processo de avaliação de qualidade, baseados nas normas de Qualidade ISO/IEC 9126⁵, num conjunto de características presentes em Pressman (2001), Lima *et al.* (2000) in Moraes e Werneck além de Reis *et al.* (2002), Trochim (1999).

C o n s t r u t o	
Classes de Valorização	CrITÉRIOS de Qualidade
Usabilidade	Entendimento global do <i>site</i> , Feedback e help on-line, Características estéticas da interface e Características especiais
Funcionalidade	Capacidade de busca e recuperação, Características de navegação e navegador e Características relacionadas ao domínio da aplicação
Confiabilidade	Processamento correto de links, Recuperação de erros e Validação e recuperação do input do usuário
Eficiência	Desempenho do tempo de resposta, Velocidade de geração da página e Velocidade da geração de gráficos
Manutenibilidade	Facilidade de correção, Adaptabilidade e Extensibilidade

Quadro 10 – Tipos de classes de valorização e critérios de qualidade na construção dos construtos.

Fonte: Moraes e Werneck, 2000

2.4 QI EM SITES E O PROCESSO COGNITIVO

Os trabalhos apresentados anteriormente enumeram uma série de dimensões através das quais a QI pode ser avaliada. Diversas destas dimensões apresentam características técnicas, para as quais indicadores ou índices numéricos podem ser estabelecidos.

Entretanto, o valor da qualidade ainda assim tem, no processo cognitivo de quem utiliza a informação, grande importância. Estes processos são de difícil tratamento analítico. Uma proposta para o tratamento desta questão é fornecida pela Teoria dos Conjuntos Difusos.

Para identificar a importância relativa das características individuais das diferentes dimensões da Qualidade da Informação, uma classe especial de Regressão denominada Regressão Logit pode ser utilizada.

A análise e observação das ferramentas apresentadas pelos autores, em suas diversas abordagens, considerando aspectos funcionais, metodológicos e temporais, permitem que se estabeleça um quadro apontando os principais elementos da análise da qualidade. O Quadro abaixo fornece uma sumarização das dimensões mais comuns e a frequência com que elas estão incluídas nas ferramentas descritas⁶.

⁶ Resumo realizado a partir do artigo de Knight e Burn (2005). Cada dimensão da qualidade é também descrita de forma sumária.

Ordem	Dimensão	Frequência	Definição da Dimensão
1	Precisão	8	Dados corretos, confiáveis e livres de erros
2	Consistência	7	Informação apresentada no mesmo formato e compatível com dados anteriores
3	Segurança	7	Acesso a informação restrito de forma adequada para manter a segurança
4	Atualidade	7	Atualizada para a tarefa que se tem que realizar
5	Compleitude	5	Informação não ausente, suficientemente ampla e profunda para a tarefa que se tem que realizar
6	Concisão	5	Informação representada compactamente sem ser rebuscada demais
7	Confiabilidade	5	Informação correta e confiável
8	Acessibilidade	4	Informação disponível ou fácil e rapidamente recuperável
9	Disponibilidade	4	Informação fisicamente acessível
10	Objetividade	4	Informação não viesada, sem preconceitos ou imparcial
11	Relevância	4	Informação importante e útil para a tarefa que se tem que realizar
12	Usabilidade	4	Informação clara e de fácil utilização
13	Compreensibilidade	4	Dados claros, sem ambigüidade e facilmente compreensíveis
14	Quantidade de dados	3	Quantidade ou volume de dados disponíveis apropriado
15	Credibilidade	3	Informação classificada como crível e verdadeira
16	Navegabilidade	3	Dados facilmente encontrados e ligações adequadas
17	Reputabilidade	3	Informação altamente classificada em termos de fonte ou conteúdo
18	Utilidade	3	Informação aplicável e útil para a tarefa que se tem que realizar
19	Eficiência	3	Dados capazes de rapidamente encontrar a informação necessária para a tarefa que se tem que realizar
20	Valor agregado	3	Informação benéfica e oferece vantagem com o uso

Quadro 11 – Tipos de dimensão, frequências e definição da dimensão.

Fonte: Knight e Burn, 2005

Neste trabalho apresenta-se uma metodologia IQM (*Information Quality Measuring*) para assinalar critérios de qualidade com as ferramentas de mensuração adequadas e as dimensões agrupadas em categorias por similaridade e pertinência.

3 CONJUNTOS DIFUSOS

A Teoria dos Conjuntos Difusos (*Fuzzy Set Theory*) foi introduzida por Zadeh (1965) para resolver problemas baseados em dados expressos de forma qualitativa ou afetados por incerteza.

3.1 INTRODUÇÃO

Zadeh (1974) defende que: “embora as técnicas matemáticas convencionais tenham sido e continuarão a ser aplicadas à análise de sistemas humanísticos, está claro que a grande complexidade de tais sistemas exige abordagens que sejam significativamente diferentes dos métodos tradicionais - métodos que são altamente eficientes quando aplicados a sistemas mecânicos, mas deixam de ser precisos em relação aos sistemas nos quais o comportamento humano tem um papel importante.” (GONÇALVES, 2005)

A crescente disciplina da lógica difusa fornece por si mesma um meio para conectar estas tarefas. Em determinado nível, a lógica difusa pode ser vista como uma linguagem que permite transferir sentenças sofisticadas da linguagem natural a uma linguagem matemática formal.

A Teoria de Conjuntos Difusos integra um conjunto de conceitos, operadores, regras e princípios que constituem uma extensão de diversos domínios e teorias clássicas.

A abordagem difusa tem em conta o grau de incerteza por vezes inerente a características e a classificações, usando a teoria dos conjuntos difusos para modelar esse grau de incerteza.

No presente trabalho, será utilizada a abordagem difusa para estabelecer padrões de qualidade em *sites* na *web*.

3.2 CARACTERÍSTICAS GERAIS DOS CONJUNTOS DIFUSOS

De modo geral, a teoria dos conjuntos difusos permite que elementos distintos possuam graus de pertinência a vários conjuntos, de modo a tornar possível a representação matemática de conceitos vagos e imprecisos.

O grau de pertinência de um elemento de um universo a um determinado conjunto difuso é representado por um número real no intervalo $[0,1]$, que representa o quão verdadeira é a afirmação de que esse elemento pertence àquele conjunto.

O conjunto difuso A em X pode ser definido como o conjunto de pares ordenados tais que:

$$A = \{[x, \mu_A(x)] | x \in X\} \quad \text{Equação 2}$$

onde $\mu_A(x)$ é a função de pertinência ao conjunto difuso.



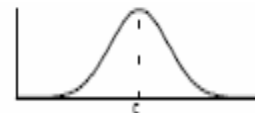

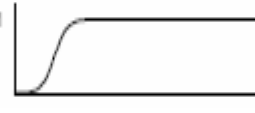
As funções de pertinência podem ser do tipo discreto ou contínuo. A notação em cada caso é dada por:

$$\text{a. No caso discreto: } A = \sum_U \mu_A(U) / U \quad \text{Função 1}$$

$$\text{b. No caso contínuo: } A = \int_U \mu_A(U) / U \quad \text{Função 2}$$

onde os símbolos Σ e \int não têm representação algébrica.

Existem diferentes alternativas de relações de pertinência sendo que as mais comumente usadas, no caso contínuo, são triangular, trapezoidal, de Gauss, de Bell e sigmoidal, cujas fórmulas e gráficos estão apresentados no Quadro 12.

Função	Fórmula e Parâmetros	Gráfico
Triangular	$\mu_A(X) = \max \left\{ \min \left[\frac{(x-a)}{(b-a)}; \frac{(c-x)}{(c-b)} \right]; 0 \right\}$	
Trapezoidal	$\mu_A(X) = \max \left\{ \min \left[\frac{(x-a)}{(b-a)}; 1; \frac{(d-x)}{(d-c)} \right]; 0 \right\}$	
Gaussiana	$\mu_A(X) = e^{-\frac{1}{2}[(x-c)/\sigma]^2}$	
Bell	$\mu_A(X) = \left[1 + \left \frac{(x-c)}{a} \right ^{2b} \right]^{-1}$	
Sigmoidal	$\mu_A(X) = \left[1 + \exp(-a(x-c)) \right]^{-1}$	

Quadro 12 – Tipos de relações de pertinência contínuas
Fonte: Pacheco, 1996

A teoria de conjuntos difusos possui três operações básicas ou operadores padrão, originalmente definidas por Zadeh (1965) que são apresentadas a seguir.

3.3 OPERADORES PADRÃO

Essa seção é baseada em Durk (1994), que trata apenas dos operadores padrão da teoria dos conjuntos difusos, então as referências a operadores difusos aqui devem ser consideradas relativamente aos operadores padrão.

Para ilustrar este efeito, pode-se considerar uma observação sobre as alturas de pessoas. Uma pessoa de 1,85 m pode ser considerada alta em um grupo de pessoas altas com 0,7 de pertinência.

3.3.1 União

A teoria de conjuntos, na teoria clássica de conjuntos, afirma que a união de conjuntos é formada pelo conjunto dos elementos distintos que pertencem a pelo menos um deles ou a todos simultaneamente. Assim sendo, os elementos do conjunto união, não podem possuir valor de pertinência menor do que a possuída em qualquer um dos conjuntos originais. A forma padrão que a lógica difusa usa para obter a intersecção entre dois conjuntos é a seguinte:

$$\begin{aligned}\mu_{A \cup B}(x) &= \max[\mu_A(x), \mu_B(x)] \text{ para todo } x \in X \\ &= \mu_A(x) \cup \mu_B(x) \\ &= \mu_A(x) \cup \mu_B(x)\end{aligned}$$

O símbolo \cup (chamado "ou" lógico) é usado na lógica difusa para representar a operação "max", que toma o valor máximo dentre os valores em consideração. Exemplo:

$$\begin{aligned}\text{ALTO} &= (0/5, 0,2/5,5, 0,5/6, 0,8/6,5, 1/7) \\ \text{BAIXO} &= (1/5, 0,8/5,5, 0,5/6, 0,2/6,5, 0/7) \\ \mu_{\text{ALTO} \cup \text{BAIXO}}(x) &= (1/5, 0,8/5,5, 0,5/6, 0,8/6,5, 1/7)\end{aligned}$$

A união entre os conjuntos dados por ALTO e BAIXO, pode ser representada por $(\mu_{\text{ALTO}} \cup \mu_{\text{BAIXO}})(x)$ e pode ser entendida como as pessoas que não são de estatura média, por exemplo, e também usada como um novo conjunto difuso.

3.3.2 Intersecção

O conjunto intersecção entre conjuntos, é aquele que contém todos os elementos que são comuns a todos. Todavia esta abordagem se diferencia na teoria dos difusos, na qual, o elemento pode pertencer parcialmente a dois ou mais conjuntos, ainda que não pertença completamente a nenhum deles. Assim, quando é considerada a intersecção desses conjuntos, não se pode dizer que um elemento possa pertencer mais ao conjunto da intersecção do que a qualquer um dos conjuntos originais.

De acordo com isto, o operador difuso usado para gerar a intersecção entre dois conjuntos difusos A e B definidos em X é dada por:

$$\begin{aligned}\mu_{A \cap B}(X) &= \min[\mu_A(x), \mu_B(x)] \text{ para todo } x \in X \\ &= \mu_A(x) \cap \mu_B(x) \\ &= \mu_A(x) \frown \mu_B(x)\end{aligned}$$

O símbolo \cap ("e" lógico) é usado para representar o operador "min", que simplesmente toma o mínimo entre os valores em consideração. Exemplo:

$$\begin{aligned}\text{ALTO} &= (0/5, 0,2/5,5, 0,5/6, 0,8/6,5, 1/7) \\ \text{BAIXO} &= (1/5, 0,8/5,5, 0,5/6, 0,2/6,5, 0/7) \\ \mu_{\text{ALTO}} \cap \mu_{\text{BAIXO}}(x) &= (0/5, 0,2/5,5, 0,5/6, 0,2/6,5, 0/7)\end{aligned}$$

A intersecção entre os conjuntos dados por ALTO e BAIXO, dada por $(\mu_{\text{ALTO}} \cap \mu_{\text{BAIXO}})(x)$, pode ser entendida como as pessoas de média estatura, por exemplo, e usada como um novo conjunto difuso.

3.3.3 Complementação

O conjunto complementar é aquele que possui todos os elementos pertencentes ao domínio, mas que não fazem parte de um dado conjunto A. Definido um conjunto difuso A, é possível encontrar o seu complemento \bar{A} através da seguinte expressão: $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$.

Por exemplo, dado o referido conjunto das *peessoas altas*, usado nos exemplos anteriores, este operador pode ser usado para criar o conjunto das *peessoas não altas* ou *peessoas médias ou baixas*.

$$\mu_A(x) = \text{ALTO} = (0/5, 0,2/5,5, 0,5/6, 0,8/6,5, 1/7)$$

$$\mu_{\bar{A}}(x) = \text{NÃO ALTO} = (1/5, 0,8/5,5, 0,5/6, 0,2/6,5, 0/7)$$

O complementar de ALTO, NÃO ALTO e BAIXO, dada por $\mu_{\bar{A}}(x)$, pode ser entendida como as pessoas de baixa e média estatura, por exemplo, e usada como um novo conjunto difuso.

3.4 PROBLEMAS DE OTIMIZAÇÃO DIFUSA

Os problemas de otimização difusa são introduzidos por Bellman, Zadeh e Zimmerman citados em PEDRYCZ, GOMIDE, 1998 além Zemankova e Kandel (1984) Zhang et al (1983) nos quais tanto as funções objetivo quanto as restrições são situações mal definidas e representadas por conjuntos difusos.

Seja X um espaço de decisão, uma função de pertinência $D(x)$ é representada por:

$$D(x) = F_k(x) * C_l(x), \quad k = 1, \dots, N \quad \text{e} \quad l = 1, \dots, M \quad \text{Função 3}$$

onde:

- F_k são funções objetivo difusas;
 $C_l(\mathbf{x})$ são as restrições definidas em \mathbf{X} ;
 * é o operador de agregação adequado.

Se existir um subconjunto $\mathbf{x} \subseteq \mathbf{X}$ para o qual $D(\mathbf{x})$ atinge um máximo, então \mathbf{X} é dito ser decisão máxima (se o problema de otimização for de maximização). Se $\mathbf{x} \in \mathbf{X}$ resolve simultaneamente as funções objetivo e restrições com maior grau de pertinência, então determinar este conjunto \mathbf{x} é equivalente a resolver o problema de otimização:

$$\begin{aligned} &\max \lambda \\ &\text{sujeito a } F_k(x) \geq \lambda, k = 1, \dots, N \\ &\quad C_l(x) \geq \lambda, l = 1, \dots, M \quad \text{Função 4} \\ &\lambda \in (0,1]; x \in X \end{aligned}$$

3.5 PROBLEMAS DE OTIMIZAÇÃO COM FUNÇÕES E RESTRIÇÕES DIFUSAS

O modelo de otimização associado com programação linear para o qual as funções objetivo e restrições são difusas é dado por:

$$\begin{aligned} &\max_f c^T x \\ &\text{sujeito a } Ax \leq_f b \quad \text{Função 5} \\ &x \geq 0 \end{aligned}$$

onde \max_f significa que a função objetiva é difusa e que a desigualdade \leq_f é compreendida como essencialmente menor ou igual a, sendo a i -ésima restrição associada a $Ax \leq_f b$ por $(Ax) \leq_f b_i$, $i = 1, \dots, M$. Neste caso é assumido que c e A , bem como um nível de aspiração b_0 e uma tolerância

máxima p_i para as restrições, são previamente estabelecidos. Neste caso o problema passa a ser:

$$\begin{aligned} c^T x &\geq_f b_0 \\ Ax &\leq_f b \\ x &\geq 0 \end{aligned} \quad \text{Função 6}$$

com função de pertinência dada por:

$$F(x) = \begin{cases} 1, & \text{se } c^T x > b_0 \\ 1 - \left[\frac{b_0 - c^T x}{p_0} \right], & \text{se } b_0 - p_0 \leq c^T x \leq b_0 \\ 0, & \text{se } c^T x < b_0 - p_0 \end{cases}$$

Função 7

$$C_i(x) = \begin{cases} 1, & \text{se } (Ax)_i < b_i \\ 1 - \left[\frac{(Ax)_i - b_i}{p_i} \right], & \text{se } b_i \leq (Ax)_i \leq b_i + p_i \\ 0, & \text{se } (Ax)_i > b_i + p_i \end{cases}$$

o Gráfico 1 mostra a função de pertinência $F(x)$ para um nível de aspiração igual a 0,8 e um nível de tolerância igual a 0,5.

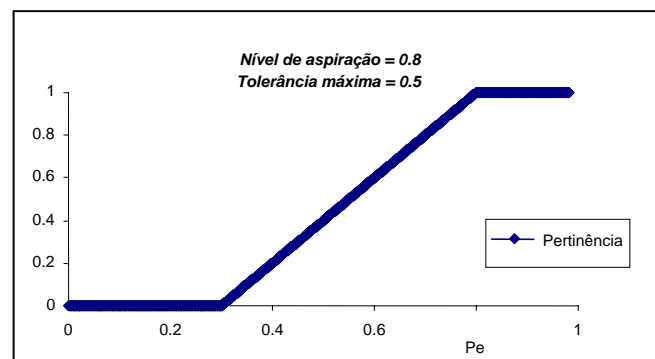


Gráfico 1 – Gráfico da função de pertinência

A Função 3, p.45, pode ser traduzido pelo seguinte problema:

$$\begin{array}{ll}
 \min \theta & \\
 \text{sujeito a } c^T x \geq b_0 - \theta p_0 & \\
 (Ax)_i \leq b_i + \theta p_i, i = 1, \dots, m & \text{Função 8} \\
 \theta \in (0, 1]; x \geq 0 &
 \end{array}$$

A solução deste problema de otimização vai permitir a decisão da aplicação no grupo de dimensão da qualidade que melhor representa a aspiração dos usuários quanto à percepção da qualidade no *site*.

4 A REGRESSÃO

O termo regressão foi empregado pela primeira vez por Francis Galton (1822-1911) num estudo da relação entre as alturas dos pais e filhos.

4.1 INTRODUÇÃO

A análise de regressão é uma das técnicas mais utilizadas para analisar dados (CHATTERJEE e PRICE, 1991). O mote principal dos modelos de regressão é explicar uma ou várias variáveis de interesse (objetivo) em função de outras variáveis (explicativas). Dentro da estatística, muitos métodos de análise utilizam a teoria de regressão.

4.2 REGRESSÃO LOGÍSTICA - *LOGIT*

Esta teoria pode ser encontrada em Neter, Kutner, Nachtsheim e Wasserman (1996) e Draper e Smith (1981), entre outros. Neste trabalho, o interesse imediato é o de aplicar um modelo específico de regressão: o modelo de regressão quando a variável resposta é do tipo binário. Estes modelos são conhecidos como regressão binária ou regressão logit.

Regressão Binária ou Logística é uma forma de regressão utilizada quando a variável dependente é dicotômica e as variáveis regressoras são de qualquer tipo. Quando a variável dependente apresenta mais de duas

classes, tem-se o caso de regressão logística multinomial e, quando estas classes podem ser classificadas, tem-se o caso de regressão logística ordinal.

As regressões logísticas têm sido utilizadas principalmente para:

- a. prever uma variável dependente com base em valores de variáveis independentes categóricas ou contínuas, explicitando a percentagem da variância da variável dependente que é explicada pelo conjunto de regressores;
- b. classificar a importância relativa do conjunto de regressores;
- c. estabelecer efeitos de interações entre os regressores utilizados;
- d. compreender o impacto das covariâncias das variáveis controle.

A técnica de regressão logística aplica o método de estimação por máxima verossimilhança depois de transformar a variável dependente em uma variável logit (o logaritmo das chances da variável dependente ocorrer ou não). Assim, a regressão logística estima a probabilidade de certo evento ocorrer ou não (no caso das técnicas de mínimos quadrados, os coeficientes das variáveis regressoras determinam mudanças diretas no valor da variável dependente).

4.3 O MODELO DE REGRESSÃO LOGÍSTICA

Considere-se uma variável Y que assuma apenas dois valores que podem ser dados por:

$$\begin{cases} P(y = 1) = \pi \\ P(y = 0) = 1 - \pi \end{cases} \quad \text{Equação 3}$$

Transformando a variável resposta em proporções de vezes em que a mesma assume o valor 0 ou o valor 1, é possível construir um modelo da seguinte forma:

$$\pi = a + b(x) \quad \text{Equação 4}$$

que é um modelo de probabilidade linear e que implica que a probabilidade de sucesso é uma função linear de X (conjunto de regressores). O modelo (x), contudo, é um modelo limitado, pois em primeiro lugar permite valores fora do intervalo $[0,1]$. Em segundo lugar, os erros não são normalmente distribuídos.

Na regressão logística, o relacionamento entre π e X (conjunto de regressores) é um relacionamento curvilíneo em vez de linear. A linearização pode ser obtida através da seguinte equação:

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta X \quad \text{Equação 5}$$

A razão $\pi/(1-\pi)$ é conhecida como razão *odds* e representa a probabilidade de mudança dado um conjunto de valores para os regressores.

A Equação 5 é chamada de *transformação logística* ou *logit*. Esta transformação assegura que os valores de π estarão compreendidos no intervalo $[0,1]$.

Quanto às formas, o modelo de regressão logística pode ser aditivo ou multiplicativo. As equações 6 e 7 mostram ambas as formas, respectivamente.

$$\log[\pi/(1-\pi)] = a + \beta X \quad \text{Equação 6}$$

$$[\pi/(1-\pi)] = \exp(a) * \exp(\beta X) \quad \text{Equação 7}$$

As principais funções de ligação são mostradas no Quadro 13.

Função	Função de Ligação	Distribuição
Logit	$\log[\pi/(1-\pi)]$	Logística
Probit	$\theta^{-1}(\pi)$	Normal
Gompit	$\log[-\log(1-\pi)]$	Gompertz

Quadro 13 – Funções de Ligação – Distribuição Logística

4.4 CRITÉRIOS PARA AVALIAR AJUSTE DE UM MODELO

A análise do ajuste de um modelo de regressão logística é realizada com base na avaliação da função de log verossimilhança. Quanto mais próximo de zero for o valor de -2 log verossimilhança, melhor o ajuste encontrado. Diferenças nos valores obtidos determinam melhorias no ajuste global. A estatística do teste é baseada em uma distribuição qui-quadrado.

Outras medidas para avaliar o ajuste são:

Desvios (Deviance)

$$\text{Estatística: } D = 2 \sum O_j \log \left(\frac{O_j}{E_j} \right) \quad \text{Equação 8}$$

Interpretação: O modelo procura o menor conjunto de desvios entre valores observados e estimados.

Teste de Hosmer e Lemeshow

Estatística: Cria percentuais de risco.

Interpretação: Utiliza teste do qui-quadrado para comparar valores observados e estimados. Mais indicado quando os dados possuem diferentes padrões de covariância.

Estatística Wald

$$\text{Estatística: } W = \frac{(\hat{\theta} - \theta)^2}{\text{var}(\hat{\theta})} \quad \text{Equação 9}$$

Interpretação: Utilizado para verificar a presença ou não de um efeito. A estimativa de máxima verossimilhança de um parâmetro de interesse é comparada com o valor proposto para o parâmetro, com a suposição de que a diferença entre eles será aproximadamente normal. O teste utiliza a comparação do quadrado da diferença com uma distribuição qui-quadrado.

Teste Escore

$$\text{Estatística: } \left(\frac{\partial \log L(\theta | x)}{\partial \theta} \right)_{\theta=\theta_0} \geq C \quad \text{Equação 10}$$

onde L representa a função de verossimilhança de θ_0 . e C representa uma constante arbitrária.

Interpretação: O teste escore é um teste estatístico utilizado para avaliar uma hipótese nula simples (por exemplo, que o parâmetro de interesse θ é igual a algum valor particular).

4.5 INTERPRETAÇÃO DOS COEFICIENTES

O valor β_j representa as alterações no valor de logit para uma mudança unitária no valor do regressor X_j . O valor dos *odds* é dado pela expressão: $OR = \exp(\beta_j)$ assim, o percentual de mudança no valor das razões *odds* pode ser obtido através de $100 \cdot (OR - 1)$.

Algumas questões relevantes são:

- a. Considere uma amostra aleatória simples: ajustes teriam de ser realizados para dados *clusterizados*, estudos de caso com variáveis de controle, etc.;
- b. Para comparar modelos aninhados, a amostra deve ser a mesma. Poderia haver um problema ocasionado por valores faltantes;
- c. Existem diversas estimações do ajustamento; a estatística de desvios é útil quando todos os preditores são categóricos; ajustes podem ser realizados quando não for o caso;
- d. Problemas surgem quando os dados são insuficientes ou esparsos;
- e. Com dados esparsos, a convergência na estimação de parâmetros se torna difícil, os coeficientes b's são superestimados, e o erro-padrão pode ser enorme ou inexistir;
- f. Um valor de colinearidade acima de 0,80 ou um modelo excessivamente complexo pode gerar problemas adicionais de estimação.

4.6 REGRESSÃO LOGÍSTICA E O PACOTE SPSS

Os comandos e procedimentos para realizar regressão logística utilizando o pacote SPSS são dados por:

► From the menus choose:

Analyze

Regression

Binary Logistic...

- ▶ Select one dichotomous dependent variable. This variable may be numeric or short string.
- ▶ Select one or more covariates. To include interaction terms, select all of the variables involved in the interaction and then select **>a*b>**.

To enter variables in groups (**blocks**), select the covariates for a block, and click **Next** to specify a new block. Repeat until all blocks have been specified.

Optionally, you can select cases for analysis. Click **Select**, choose a selection variable, and click **Rule**.

A seguir, são mostrados os principais resultados obtidos no processamento de regressão binária (ou logística). Estes resultados são transcritos do *site* www.spss.com.br.

Os primeiros resultados referem-se a casos que foram incluídos na análise, códigos utilizados para as variáveis dependentes e códigos para quaisquer variáveis categóricas listadas.

Case Processing Summary			
Unweighted Cases ^a		N ^b	Percent ^c
Selected Cases	Included in Analysis ^d	200	100.0
	Missing Cases ^e	0	.0
	Total ^f	200	100.0
Unselected Cases ^g		0	.0
Total ^h		200	100.0

a. If weight is in effect, see classification table for the total number of cases.

a. Unweighted Cases –

b. N – Número de casos em cada categoria.

c. Percent – Percentual de casos em cada categoria.

d. Included in Analysis – Esta linha fornece o número e o percentual de casos que foram incluídos na análise. Quando não existirem valores *missing*, será igual ao número de casos.

- e. **Missing Cases** – Fornece o número e o percentual de casos *missing*. O procedimento *logistic regression* do SPSS exclui os casos com valores ausentes.
- f. **Total** – Total de casos utilizados na análise incluindo os casos *missing*.
- g. **Unselected Cases** – Casos não selecionados.
- h. **Total** – Total de casos utilizados na análise incluindo os casos *missing* e os **Unselected Cases**.

Dependent Variable Encoding

Original Value ⁱ	Internal Value ^j
.00	0
1.00	1

Categorical Variables Codings^a

		Frequency ^k	Parameter coding ^l	
			(1)	(2)
ses	1.00	47	1.000	.000
	2.00	95	.000	1.000
	3.00	58	.000	.000

^a. This coding results in indicator coefficients.

DESCRIÇÃO DO MODELO NULO

Block 0: Beginning Block

Modelo sem preditores e com apenas o intercepto. As variáveis não incluídas são descritas na tabela "*Variables not in the Equation*".

Classification Table^{a,b}

Observed ^d			Predicted ^e		
			honcomp		Percentage Correct
			.00	1.00	
Step 0 ^c	honcomp	.00	147	0	100.0
		1.00	53	0	.0
Overall Percentage ^f					73.5

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

	B ^g	S.E. ^h	Wald ⁱ	df ^j	Sig. ⁱ	Exp(B) ^k
Step 0 Constant	-1.020	.160	40.540	1	.000	.361

Variables not in the Equation

	Score ^l	df ^m	Sig. ^l
Step 0 Variables			
read	47.906	1	.000
science	34.862	1	.000
ses	14.783	2	.001
ses(1)	.302	1	.582
ses(2)	8.666	1	.003
Overall Statistics ⁿ	58.644	4	.000

c. **Step 0** – SPSS possibilita diversos passos para construir o modelo logístico. A diferença entre os passos está nos preditores que são incluídos no modelo. Este aspecto é similar a incluir uma variável por vez. O padrão do SPSS é realizar o procedimento em dois passos. O primeiro, STEP 0, não inclui nenhum preditor (apenas o intercepto).

d. **Observed** – Indica o número de 0s e 1s observados na variável dependente.

e. **Predicted** – Indica o número de 0s e 1s preditos para a variável dependente.

f. **Overall Percentage** – Percentual de casos preditos corretamente para a variável dependente para o modelo sob análise. No exemplo acima, corresponde a $73,5 = 147/200$.

g. **B** – Coeficiente para o intercepto.

h. **S.E.** – Erro padrão para o coeficiente.

i. **Wald e Sig.** – Teste Qui-Quadrado de Wald. Testa a hipótese nula de que o coeficiente é igual a zero. No exemplo, esta hipótese é rejeitada porque o valor de *p-level* (que aparece na coluna chamada **Sig.**) é menor do que um valor crítico de p (0,05 ou 0,01). Assim, para este exemplo, pode-se concluir que o valor do intercepto não é zero. Este resultado normalmente não interessa ao pesquisador.

Teste Wald: Teste utilizado para verificar a presença ou não de um efeito. A estimativa de máxima verossimilhança de um parâmetro de interesse é comparada com o valor proposto para o parâmetro, com a suposição de que a diferença entre eles será aproximadamente normal. O teste utiliza a comparação do quadrado da diferença com uma distribuição qui-quadrado. No caso univariado, a estatística do teste é dada por:

$$(\hat{\theta} - \theta)^2 / \text{var}(\hat{\theta})$$

Alternativamente pode ser utilizada a distribuição normal como base do teste. Neste caso, a estatística passa a ser:

$$(\hat{\theta} - \theta) / \text{se}(\hat{\theta})$$

onde $\text{se}(\hat{\theta})$ é uma estimativa de máxima verossimilhança para o erro padrão.

Normalmente o teste de Wald fornece os mesmos resultados do teste da razão de verossimilhança.

j. **df** – Graus de liberdade para o teste Wald. No exemplo é igual a 1 porque existe apenas um preditor no modelo (o intercepto).

k. **Exp(B)** – Razão Odds. É um resultado padrão por que razões Odd são mais facilmente interpretáveis do que o valor do coeficiente (cujo valor está em unidades *log-odds*). A chance (odds) é igual a $53/147 = 0.361$.

l. **Score e Sig.** – O teste Score é utilizado para prever se uma variável independente poderia ser significativa no modelo ou não. Verificando o valor *p*, nota-se que excetuando o primeiro valor dummy para a variável *ses* todos os demais preditores poderiam ser incluídos.

O teste score é um teste estatístico utilizado para avaliar uma hipótese nula simples (por exemplo, que o parâmetro de interesse θ é igual a algum valor particular):

$$\left(\frac{\partial \log L(\theta | x)}{\partial \theta} \right)_{\theta=\theta_0} \geq C$$

onde L representa a função de verossimilhança de θ_0 e C representa uma constante arbitrária.

m. **df** – Indica o número de graus de liberdade para cada variável.

n. **Overall Statistics** – Indica o resultado geral da inclusão de todos os preditores no modelo.

5 O MODELO PROPOSTO

5.1 INTRODUÇÃO

Para avaliar a qualidade de um *site*, deve-se compreender o usuário sob as perspectivas sociológicas, cognitivas e de como este realiza seus procedimentos de escolha. Ou seja, compreender o que motiva o usuário é imperativo para que se possa “modelar” seus conceitos de qualidade de informação dentro de um contexto e saber como podem ser estabelecidos.

Assim, é possível definir que qualidade da informação depende de três fatores principais:

- a. situação no momento;
- b. como o usuário gostaria que efetivamente fosse;
- c. sentimento global sobre a realidade.

A situação no momento é influenciada por um conjunto de dimensões de qualidade que podem ser medidas direta ou indiretamente. Independentemente de condições de acesso a tecnologia, é factível recuperar a percepção subjetiva que o usuário final tem a respeito da qualidade da informação que lhe é disponibilizada.

Por outro lado, a percepção geral do ambiente também deve ser analisada e esta percepção é função das dimensões de qualidade da informação apropriadas para um determinado *site*. Identificar quais são estas dimensões de qualidade da informação está diretamente associado ao contexto em análise.

As dimensões a serem estabelecidas podem ser obtidas diretamente da literatura de Qualidade da Informação. Por exemplo, Strong, Lee e Wang (1997) estabelecem um conjunto de dezesseis dimensões que estão agrupadas em quatro categorias.

A qualidade da informação é estabelecida a partir do ponto de vista do usuário final. Então, para avaliar a qualidade de um *site* é necessário compreender o processo de valoração das diferentes dimensões da qualidade e, também, da métrica de valoração destas dimensões pelo usuário.

5.2 OBTENDO DIMENSÕES DA QUALIDADE RELEVANTES EM UM *SITE*

Considere-se um *site* genérico, no qual estão presentes as seguintes dimensões de qualidade da informação:

- I. Acessibilidade (medida em termos de links em funcionamento)
- II. Consistência (medida em número de páginas com mesmo estilo)
- III. Concisão (medida em números de acessos em profundidade, dentro do *site*)
- IV. Clareza (medida em uma classificação própria do usuário)
- V. Compreensividade (medida em uma classificação própria do usuário)
- VI. Correção (medida em uma classificação própria do usuário)

Sem perda de generalidade, as dimensões I, II e III também podem ser medidas em uma classificação própria do usuário, desde que consistentemente aplicada esta escala de medidas. O usuário pode, ainda, avaliar o grau de importância com que gostaria de ver valorizadas estas

dimensões no *site*. Tendo realizado estes dois processos subjetivos de avaliação, o usuário pode, ainda, emitir um juízo global sobre o *site*, expressando nesta avaliação sua percepção global sobre a qualidade do mesmo.

Assim, considere-se genericamente:

- a. um conjunto de dimensões de qualidade da informação, d_i ;
- b. uma métrica comum única estabelecida para cada d_i .

Estes dois pontos podem revelar a situação do *site*, no momento em que o usuário final o acessa.

Porém, para cada uma das dimensões da qualidade da informação, o usuário pode ter uma percepção sobre a situação ideal daquela dimensão. Em outras palavras, como ele, o usuário, gostaria que fosse dentro do *site*, aquela dimensão. Então, é possível:

- c. uma métrica estabelecida para avaliar como o usuário gostaria que fosse tratada a dimensão da qualidade d_i , dentro do *site*.

Finalmente, a avaliação global é resultante do processo cognitivo e da valoração que o usuário faz a respeito das dimensões de qualidade da informação presentes. Então, pode-se obter:

- d. uma avaliação global sobre a qualidade do *site* percebida pelo usuário.

Esta situação é descrita na Figura 1, onde as dimensões, a métrica adotada no trabalho, a importância dada à dimensão e a percepção global da Qualidade Percebida estão como apresentado no instrumento de coleta de dados.

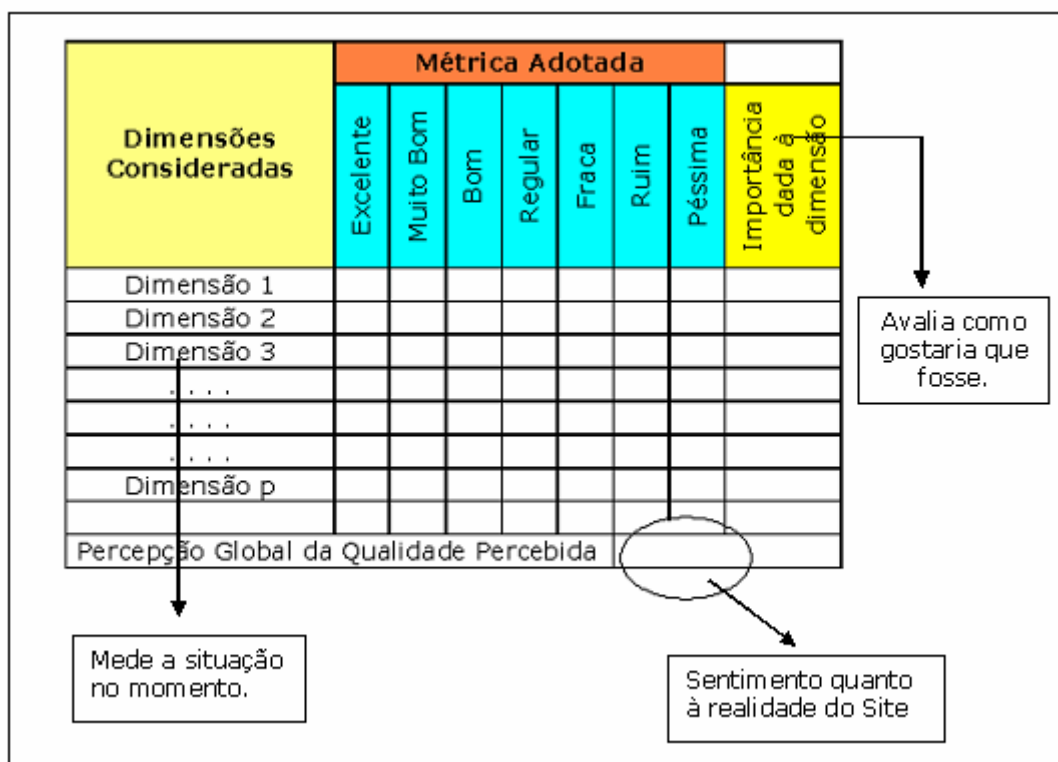


Figura 1 - Mecanismo para recuperar informação sobre a qualidade do *site*

De forma generalizada, considere-se que para o *site* que está sendo avaliado existam L dimensões sendo avaliadas e que estas dimensões possam ser agrupadas por grupos de dimensões. Por exemplo, dimensões tais como precisão e disponibilidade estão sendo consideradas dimensões pertencentes ao grupo das que avaliam a qualidade intrínseca da informação.

Neste caso, é possível criar grupos de dimensões de qualidade da informação que permitam aferir para cada grupo a qualidade total percebida pelo usuário do *site*. Então, o que se tem são:

I: grupos de dimensões

L: dimensões avaliadas dentro do Grupo I

O algoritmo descrito abaixo, pode ser calculado para determinar o valor atribuído pelos usuários para cada grupo de dimensões.

Para $i = 1, \dots, K$

para $j = 1, \dots, N$

faça

$$Cat_Dom[i, j] = \sum_{l=1}^M \frac{Q[i, l, j] * V[i, l, j]}{M}, \quad l = 1, \dots, M; \quad i = 1, \dots, K; \quad j = 1, \dots, N$$

onde:

i : i -ésimo grupo de dimensões de qualidade;

K : total de grupos de dimensões de qualidade identificados;

j : j -ésimo indivíduo pesquisado;

M : número de dimensões de qualidade incluídas no grupo i ;

N : total de indivíduos considerados;

$Q[i, l, j]$ = valor atribuído no Site para a dimensão l que está identificada no i -ésimo grupo pelo j -ésimo indivíduo;

$V[i, l, j]$ = importância dada pelo usuário à dimensão l que está identificada no i -ésimo grupo pelo j -ésimo indivíduo;

$Cat_Dom[i, j]$ = valor ponderado para o i -ésimo grupo pelo indivíduo j .

Algoritmo 1 – Valoração dos grupos de dimensões

Definindo-se uma escala arbitrária E , pode-se realizar a seguinte transformação em $Cat_Dom[i, j]$:

$$TransfE[i, j] = \frac{E * Cat_Dom[i, j]}{E * k} \quad \text{Equação 11}$$

onde k é o número de intervalos nos quais as dimensões de qualidade foram avaliadas e todos os grupos pertencerão ao intervalo $[0, E]$.

Esta variável transformada pode ainda ser separada em categorias de interesse pelos gestores do ambiente que está sendo analisado. A idéia central é a de que o tratamento por níveis pode ser mais eficiente do que

admitindo-se uma escala intervalar. A Função 9 dá exemplos de níveis que podem ser adotados.

$$f(N - E) = \begin{cases} \text{nível 1 se } TransfE[i, j] \leq n_1 \\ \text{nível 2 se } n_1 < TransfE[i, j] \leq n_2 \\ : \\ : \\ \text{nível } p \text{ se } TransfE[i, j] \geq n_p \end{cases} \quad \text{Função 9}$$

Estabelecidas as métricas e os grupos de dimensões de qualidade de interesse, é necessário estabelecer um critério para priorizar as dimensões.

5.3 DESENVOLVENDO MECANISMOS PARA PRIORIZAR DIMENSÕES

Leung (2001) sugere que qualquer iniciativa de métrica pode ser endereçada aos usuários potenciais e que estas devem ser objetivas, efetivas e informativas. Sugere ainda o seguinte *framework*:

1. Identificar o usuário;
2. Identificar a aplicação métrica, isto é, a aplicação e o processo que viabilizam o sistema;
3. Identificar as dimensões a serem consideradas;
4. Priorizar, em termos de importância, urgência etc., as dimensões a serem consideradas;
5. Desenvolver um conjunto de iniciativas que garantam a priorização das dimensões a serem consideradas.

Neste trabalho propõe-se a realização deste *framework* em etapas que considerem:

- a. as características dos indivíduos que são “clientes” do *site* e como estas podem influenciar na qualidade global percebida por eles;
- b. os grupos de dimensões, obtidas pela aplicação do Algoritmo apresentado na página 64, influentes para a qualidade percebida pelo usuário.

O item *a* está relacionado com atributos dos usuários do sistema. No item *b*, a intenção é a de estabelecer quais dimensões serão priorizadas.

5.4 IDENTIFICAÇÃO DAS CARACTERÍSTICAS E DIMENSÕES RELEVANTES POR MODELOS DE REGRESSÃO

A Função 9 pode ser utilizada também para definir intervalos sobre a qualidade global do *site* percebida pelo usuário.

Seja *QP* a qualidade global percebida pelo usuário do *site*, medida em uma escala contínua dentro de um dado intervalo. Considere-se o seguinte critério:

***Se $QP \geq \text{Limite estabelecido a priori}$,
então a qualidade do site é boa,
caso contrário, a qualidade é nula.***

Fazendo

$$QP = \begin{cases} 1 & \text{se } QP \geq L_e \\ 0 & \text{caso contrário} \end{cases} \quad \text{Equação 12}$$

como L_e é um limite arbitrado, então QP pode ser modelada utilizando-se um modelo de regressão binário.

5.4.1 Mod1 – Considerando as categorias dos usuários

O primeiro modelo está associado com as características e atributos dos indivíduos que constituem a comunidade que acessa o *site*. O segundo modelo relaciona para os indivíduos que constituem a comunidade virtual como as categorias de dimensões de qualidade afetam a valoração dada à qualidade percebida.

Mod1:

$$\log\left(\frac{QP}{1-QP}\right) = \alpha + \beta_i \text{Característica}_i, \quad i = 1, 2, \dots \quad \text{Equação 13}$$

O vetor de características pode conter informações sobre:

- sexo;
- idade;
- nível cultural;
- nível de renda;
- etc.

Este modelo informa como características pessoais e cognitivas podem interferir no processo de percepção da qualidade do *site*.

5.4.2 Mod2 - Considerando os grupos de dimensões

Mod2:

$$\log\left(\frac{QP}{1-QP}\right) = \alpha + \beta_i G_i, \quad i = 1, 2, \dots \quad \text{Equação 14}$$

onde G_i constitui-se nos diversos grupos de dimensões de qualidade identificados como relevantes.

5.4.3 Mod3 - Considerando os grupos de dimensões e as categorias

Mod3:

$$\log\left(\frac{QP}{1-QP}\right) = \alpha_1 + \phi_j \text{Característica}_j + \beta_i G_i, i = 1, 2, \dots; j = 1, 2, \dots \quad \text{Equação 15}$$

que é o modelo completo, no qual os regressores são atributos pessoais da comunidade virtual do *site* e as características definidas pelas dimensões de qualidade consideradas relevantes para o *site*.

5.5 DEFINIÇÃO DOS MODELOS DE OTIMIZAÇÃO POR PROGRAMAÇÃO DIFUSA

Para maximizar ou minimizar recursos disponíveis sejam financeiros ou pessoal o modelo permitirá identificar o grupo de dimensão da qualidade onde a aplicação dos recursos refletira de maneira mais eficiente na percepção de qualidade dos usuários do *Site*.

5.5.1 Otimização do processo de avaliação da qualidade

Em problemas reais, são raras as situações nas quais é possível representar completamente todas as restrições, os objetivos, bem como complexidades das interações de variáveis existentes. Em geral, o que se faz é adotar para o problema de decisão um conjunto de variáveis interrelacionadas e objetivos a serem maximizados, efetuando-se, então, julgamentos criteriosos.

A resolução dos modelos de regressão propostos irá informar os coeficientes estatisticamente significativos e que podem ser utilizados para explicar os valores atribuídos à percepção global do *site*.

Contudo, a utilização dos regressores identificados para situações nas quais existam limitações de recursos é mais restritiva. Nestes casos, o conceito de otimização é fundamental. Por exemplo, para que se possam melhorar algumas das dimensões contidas em um grupo de dimensões estatisticamente significativas pode ser que se disponham de recursos limitados. Tem-se aqui, portanto, um claro problema de decisão.

5.5.2 Construção dos modelos de otimização difusa

A otimização difusa além dos requisitos pode estar interessada tanto na maximização ou na minimização, dependendo da abordagem aplicada.

Maximização:

Seja, então, o seguinte problema de maximização:

Maximizar λ

sujeito a

$$g(x_{1i}) \geq \lambda, i = 1, 2, \dots$$

$$x_{1i} \leq Q_1 \quad \text{Função 10}$$

$$x_{2j} \leq T_2$$

$$\lambda \leq 1$$

$$x_{1i}, x_{2j} \geq 0$$

A primeira restrição corresponde aos resultados obtidos, respectivamente, pelos modelos 1, 2 e 3. Evidentemente, o modelo que

melhor explicar a percepção global da qualidade integrará o problema de maximização.

O segundo conjunto de restrições corresponde às restrições oriundas dos esforços de atuar sobre as características pessoais da comunidade virtual do *site*. Pode, por exemplo, corresponder a um esforço de marketing para ampliar o mercado em uma determinada faixa etária ou dentro de um substrato social.

O terceiro conjunto de restrições relaciona-se aos esforços de natureza técnica para melhorar a performance das dimensões de qualidade consideradas no estudo, por exemplo, os gastos que devem ser efetuados para proceder melhorias na dimensão Atualidade da informação.

A quarta restrição é feita no sentido de maximizar a probabilidade de considerar como sendo de boa qualidade a informação percebida pelo *site*. Esta restrição poderia ainda contar com um nível inferior, o qual estabeleceria a qualidade mínima aceitável para o *site*.

O problema apresentado pela Função 10 pode ser considerado de maximização linear difusa, o qual pode ser resolvido com algoritmos de programação linear simples ou de programação linear inteira.

Uma outra abordagem diz respeito a considerar tanto a função objetivo (resultado dos modelos de regressão logística) quanto as restrições (restrições visando atuar sobre características individuais ou sobre características técnicas) como sendo difusas. De fato, o valor obtido pelo modelo de regressão pode ser considerado um valor difuso, bem como as restrições consideradas para melhorar este valor. Neste caso, o modelo 7 pode ser utilizado.

Minimização:

$$\min \theta$$

$$\text{sujeito a } c^T x \geq b_0 - \theta p_0$$

$$(Ax)_i \leq b_i + \theta p_i, i = 1, \dots, m$$

$$\theta \in (0,1]; x \geq 0$$

Função 11

onde b_0 e b_i representam, respectivamente, níveis de aspiração desejáveis na função objetivo para as restrições, enquanto que p_0 e p_i representam níveis de tolerância admitidos. A forma para as funções de pertinência é a mesma dada no Capítulo 3, Função 7, p.46.

5.6 ANÁLISE DOS RESULTADOS

Diferentes aspectos ou componentes podem ser considerados ao analisar-se a percepção do usuário da qualidade em um *site*, como a motivação de uso, a frequência de acessos ou o tempo de utilização. O perfil dos usuários classificados por categoria, sexo, faixa etária, centro de estudos ou trabalho pode ser levado em conta para caracterizar a questão da percepção da qualidade.

A análise será realizada utilizando-se modelos de regressão logística, enquanto que o modelo de otimização do processo de avaliação da qualidade utiliza programação linear difusa, podendo ser resolvido por programação linear simples ou linear inteira.

A partir da aplicação dos modelos de regressão escolhe-se aquele que melhor representa a percepção de qualidade do *site* pelos usuários, representado pelos coeficientes mais significativos estatisticamente na análise.

Com estes coeficientes, constrói-se a equação de maximização ou minimização que, através do modelo de otimização difusa, permitirá definir em que grupo de dimensões investir para atingir um grau de qualidade superior.

6 APLICAÇÃO DO MODELO

6.1 CONSTRUÇÃO DO INSTRUMENTO DE COLETA DOS DADOS

Para o levantamento dos dados, foi construído um instrumento de coleta – formulário apresentado no Apêndice A. O formulário foi desenvolvido a partir das dimensões de qualidade descritas no Quadro 11, p. 38, que sintetizam os principais elementos de análise da qualidade em *sites*.

Assim, o formulário, utilizando uma escala Likert de sete pontos (*Likert seven-point scale*), foi empregado para coletar dados para as dimensões da qualidade que compõem os construtos do modelo de pesquisa. Os itens de mensuração para as dimensões e os construtos utilizados foram derivados da literatura de QI observada em *WebSites*.

Os valores relativos à percepção da qualidade, que avalia a situação do momento na visão do usuário, em cada uma das dimensões foram valorados através da escala: Excelente, Muito Bom, Bom, Regular, Ruim e Péssimo. As dimensões de qualidade foram inicialmente analisadas pelo método Delphi⁷.

Cinco profissionais da área foram convidados para avaliar e criticar os itens e realizar mudanças para eliminar itens repetitivos, itens técnicos não orientados aos usuários, questões dúbias, de resposta óbvia ou de difícil interpretação ou impossíveis de serem respondidas pelo usuário. Após três rodadas de avaliação, restaram dezenove atributos de qualidade de *web*.

⁷ **MÉTODO "DELPHI"** - Série de estudos conduzida, durante alguns anos, pela RAND Corporation (Santa Mônica, Califórnia, EUA), que recebeu a denominação de "Projeto Delphi".

Para obter uma valoração da importância dada pelo usuário a dimensões de qualidade, foi acrescentada, como pode ser visto no Anexo 1, uma coluna onde o respondente atribuiu notas de zero a dez a cada uma das dimensões. Além disso, uma pergunta foi formulada com o objetivo de aferir o sentimento do usuário quanto à realidade do *site*.

Antes de aplicar o formulário, um teste piloto foi conduzido entre 35 estudantes de graduação e cinco servidores técnico-administrativos, como pré-teste para o conteúdo e compreensão dos resultados. O teste tornou necessário reespecificar algumas expressões para tornar claro o significado de cada item do formulário sugerido durante o teste piloto. Feitas as alterações, o formulário pôde ser aplicado.

6.2 APLICAÇÃO DO INSTRUMENTO

A população-alvo está constituída pelos usuários do *site*, compostos por três categorias: alunos, servidores técnico-administrativos e professores em números aproximados de 17.000, 2.900 e 1.800, respectivamente.

Procurou-se para a realização da pesquisa tomar uma amostra representativa da população, sem o critério de constituir uma amostra aleatória proporcional considerando as três categorias, pelas dificuldades na utilização de tal critério. O tamanho da amostra conforme a fórmula apresentada na Equação 1, p. 16 resultou em:

$$n = \frac{0,5 * (1 - 0,5) * 1,645^2}{0,04^2} = 422,82 \cong 423 \quad \text{Equação 16}$$

Para tanto adotou-se um $\alpha=10\%$ e um $e=4\%$.

Os dados foram coletados numa amostra em duas etapas: a primeira no período de 12 a 14 de junho, num total de 80 formulários, e a segunda no período de 18 a 25 de agosto de 2006.

Para a aplicação do formulário na segunda etapa foram utilizados procedimentos diferenciados para Alunos e as duas outras categorias. Para Alunos, foram encaminhados via e-mail 15000 formulários. Já para Professores e S.T.A., no sentido de atingir também todo o conjunto de ensino da UFSC, os formulários foram entregues pessoalmente em praticamente todos os departamentos de ensino, num total de 360 'para professores e 580 para S.T.A.

Como era de se esperar, o retorno dos mesmos foi muito baixo. Assim a proporção respondida completamente foi de 2,6%, 16,0% e 8,1%, respectivamente.

O número de formulários corretamente preenchidos em cada estrato foi de 397, 58 e 47, respectivamente, totalizando 502 - o que superou a número mínimo estipulado. Deve ser destacado que no caso dos alunos o formulário foi encaminhado por meio eletrônico, o que pode justificar o baixo retorno, que também ocorreu nos S.T.A. Já no caso dos professores o retorno foi além do esperado.

6.3 RESULTADO PRELIMINAR

Como resultado preliminar, apresenta-se um resumo descritivo exploratório das variáveis consideradas no instrumento de coleta de dados, que podem ser divididos em três módulos, quais sejam:

- quanto às características dos usuários;
- quanto ao tipo e frequência de acesso;
- resumo da quantificação das dimensões;
- resumo da importância dada às dimensões;
- quanto ao sentimento da percepção da qualidade.

6.3.1 Quanto às características dos usuários

As características dos usuários foram divididas em: Categoria, Sexo, Idade, Centro, além da frequência e do tipo de acesso. Foram considerados também os valores atribuídos às avaliações das dimensões, a importância das dimensões e da percepção da qualidade.

Descrição dos usuários por Categoria e Sexo

Tabela 1 - Distribuição dos usuários por Categoria e Sexo

			SEXO		Total
			Feminino	Masculino	
CATEG	Aluno	Count	210	187	397
		% within CATEG	52,9%	47,1%	100,0%
		% within SEXO	81,7%	76,3%	79,1%
	Serv. Professor	Count	19	39	58
		% within CATEG	32,8%	67,2%	100,0%
		% within SEXO	7,4%	15,9%	11,6%
	Serv. Tec. Adm.	Count	28	19	47
		% within CATEG	59,6%	40,4%	100,0%
		% within SEXO	10,9%	7,8%	9,4%
Total	Count	257	245	502	
	% within CATEG	51,2%	48,8%	100,0%	
	% within SEXO	100,0%	100,0%	100,0%	

Pode ser visto que o total de alunos que compõe a amostra é de 397, enquanto que professores e técnico-administrativos representam 58 e 47 respectivamente. Os alunos na distribuição por sexo representam no feminino 210 (52,9%) e no masculino 187 (47,1%), enquanto que para

professores estes valores são dados por 19 (32,8%) e 39 (67,2%), e para técnico-administrativos, por 28 (59,6%) e 19 (40,4%), respectivamente.

Quando é observado o número de usuários pesquisados que responderam à pesquisa, pode ser visto que 79,1% são alunos, 11,6% são professores e 9,4% correspondem a servidores técnico-administrativos.

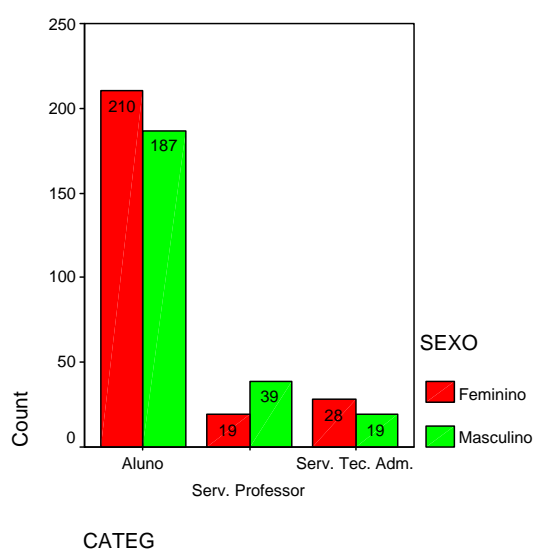


Gráfico 2 - Distribuição dos usuários por Categoria e Sexo

Descrição dos usuários por Idade

Tabela 2 - Descrição da Idade por Categoria e Sexo do usuário

IDADE		Minimum	Median	Mean	Maximum	N	Std. Deviation
Aluno	Feminino	18	22	22,9	51	208	5,321
	Masculino	17	22	23,6	52	186	6,048
Serv. Professor	Feminino	29	45	44,4	53	19	6,318
	Masculino	33	48	48,2	64	39	7,500
Serv. Tec. Adm.	Feminino	30	45	44,5	53	27	4,750
	Masculino	42	48	47,7	56	19	4,083
Total	Feminino	18	22	26,8	53	254	9,885
	Masculino	17	23	29,4	64	244	12,092

Pode ser visto que as médias das idades são sempre maiores para o sexo masculino em todas as categorias, e que a variação das idades dada pelo desvio padrão no total de usuários pesquisados é também maior para o segmento masculino: 12,092 contra 9,885 anos.

A distribuição das idades como é usual se aproxima de uma normal, sendo que a categoria cuja idade apresenta a maior assimetria é a dos alunos, como pode ser visto no Gráfico 2 (Box-Plot) abaixo e nos Gráficos 3, 4 e 5 na página a seguir.

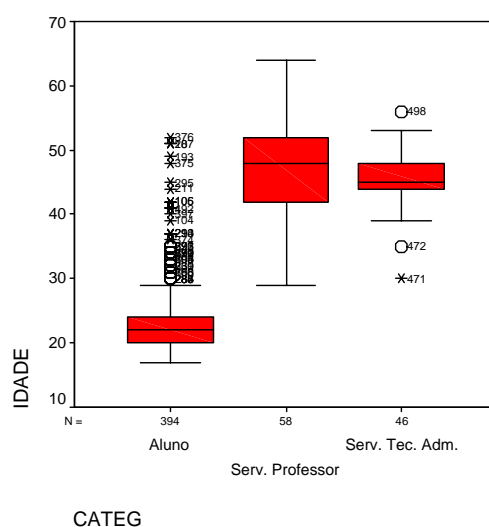


Gráfico 3 - Distribuição das Idades pelas Categorias

Os pontos acima da primeira e terceira "caixa" representam valores que, segundo o padrão estatístico, podem ser considerados extremos (próximos) ou então outliers (distantes). Neste caso estão indicando idades altas para os alunos, fora do padrão (Média=23,22 anos e Desvio Padrão=5,681), o que pode ser visto no Gráfico 4.

Descriptive Statistics^a

	N	Minimum	Maximum	Mean	Std. Deviation
IDADE	394	17	52	23,22	5,681
Valid N (listwise)	394				

a. CATEG = Aluno

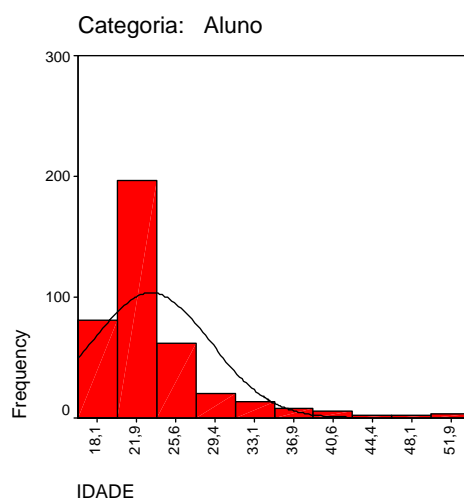


Gráfico 4 - Distribuição das Idades dos Alunos

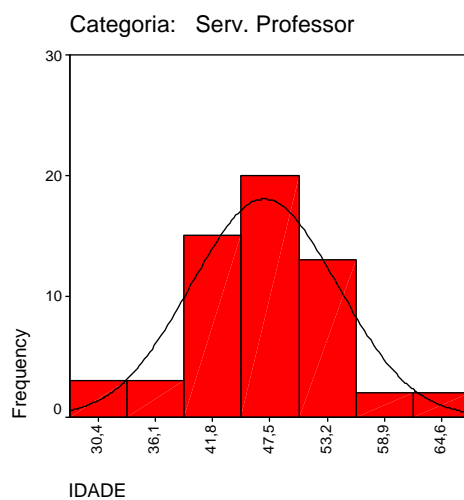


Gráfico 5 - Distribuição das Idades dos Professores

A distribuição das idades dos professores tem um comportamento aproximadamente normal centrado na Média=46,93 anos com um Desvio Padrão=7,305 anos.

Descriptive Statistics^a

	N	Minimum	Maximum	Mean	Std. Deviation
IDADE	58	29	64	46,93	7,305
Valid N (listwise)	58				

a. CATEG = Serv. Professor

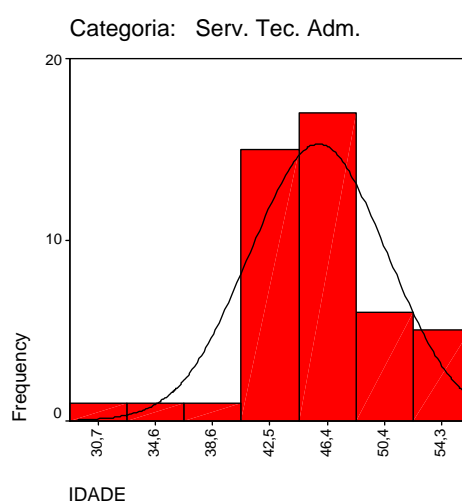


Gráfico 6 - Distribuição das Idades dos Técnico-administrativos

Já a distribuição das idades dos técnico-administrativos tem um comportamento assimétrico negativo com Média=45,83 anos com um Desvio Padrão=4,711 anos o que indica neste caso uma variação menor que a dos professores (D.P.=7,305) e também dos alunos (D.P.=5,681 anos).

Descriptive Statistics^a

	N	Minimum	Maximum	Mean	Std. Deviation
IDADE	46	30	56	45,83	4,711
Valid N (listwise)	46				

a. CATEG = Serv. Tec. Adm.

Tabela 3 - Estatísticas da variável Idade por Categoria

IDADE			
Aluno	N	Valid	394
		Missing	3
	Mean		23,22
	Median		22,00
	Mode		20
	Std. Deviation		5,681
	Skewness		2,453
	Kurtosis		7,199
	Minimum		17
	Maximum		52
	Percentiles	25	20,00
		50	22,00
		75	24,00
Serv. Professor	N	Valid	58
		Missing	0
	Mean		46,93
	Median		48,00
	Mode		48
	Std. Deviation		7,305
	Skewness		-,082
	Kurtosis		,102
	Minimum		29
	Maximum		64
	Percentiles	25	42,00
		50	48,00
		75	52,00
Serv. Tec. Adm.	N	Valid	46
		Missing	1
	Mean		45,83
	Median		45,00
	Mode		44
	Std. Deviation		4,711
	Skewness		-,611
	Kurtosis		2,208
	Minimum		30
	Maximum		56
	Percentiles	25	43,75
		50	45,00
		75	48,25

As estatísticas da variável idade são apresentadas na tabela considerando em separado as três categorias sendo que citando a assimetria (Skewness) pode-se ver que a única positiva é a relativa aos alunos 2,453 o que indica uma "cauda" mais longa à direita, enquanto que a dos professores e técnico-administrativos de -0,82 e -0,611 ambas negativas.

Descrição dos usuários por Centro de Ensino

Tabela 4 - Distribuição dos usuários por local de trabalho-estudo

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	CCA	48	9,6	9,6	9,6
	CCB	26	5,2	5,2	14,7
	CCE	29	5,8	5,8	20,5
	CCJ	10	2,0	2,0	22,5
	CCS	71	14,1	14,1	36,7
	CDS	26	5,2	5,2	41,8
	CED	21	4,2	4,2	46,0
	CFH	65	12,9	12,9	59,0
	CFM	56	11,2	11,2	70,1
	CSE	63	12,5	12,5	82,7
	CTC	68	13,5	13,5	96,2
	REIT	19	3,8	3,8	100,0
	Total	502	100,0	100,0	

Pode-se observar que todos os centro de ensino responderam à pesquisa tendo as maiores participações ocorrido nos centros CCS e CTC com 14,1% e 13,5% sendo a menor do CCJ com apenas 2,0 %.

Tipo de acesso na UFSC e na residência

- **Na UFSC:**

Tabela 5 - Tipo de acesso na UFSC

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A cabo	475	94,6	94,6	94,6
	WireLess	27	5,4	5,4	100,0
	Total	502	100,0	100,0	

Pode-se observar que 94,6 % dos usuários do *site* na UFSC o fazem através de conexão a cabo e apenas 5,4% por WireLess.

- ***Na residência:***

Tabela 6 - Tipo de acesso na residência

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Discada	130	25,9	33,4	33,4
	ADSL	259	51,6	66,6	100,0
	Total	389	77,5	100,0	
Missing	9	113	22,5		
Total		502	100,0		

Quanto ao acesso da residência chama a atenção que praticamente o dobro 66,6% possui ADSL, contra 33,4% que usam linha discada e 22,5% não declarou o tipo porque provavelmente não possui *internet* em casa.

Frequência de acessos na UFSC, por categoria

Tabela 7 - Frequência de acesso na UFSC, por Categoria

			CATEG			Total
			Aluno	Serv. Professor	Serv. Tec. Adm.	
N_UFSC	Diversas vezes ao dia	Count	56	29	34	119
		% within N_UFSC	47,1%	24,4%	28,6%	100,0%
		% within CATEG	15,3%	51,8%	72,3%	25,4%
	Praticamente todo dia	Count	60	12	11	83
		% within N_UFSC	72,3%	14,5%	13,3%	100,0%
		% within CATEG	16,4%	21,4%	23,4%	17,7%
	Três vezes por semana	Count	77	2	2	81
		% within N_UFSC	95,1%	2,5%	2,5%	100,0%
		% within CATEG	21,1%	3,6%	4,3%	17,3%
	Uma vez por semana	Count	77	7	0	84
		% within N_UFSC	91,7%	8,3%	,0%	100,0%
		% within CATEG	21,1%	12,5%	,0%	17,9%
	De 15 em 15 dias	Count	22	3	0	25
		% within N_UFSC	88,0%	12,0%	,0%	100,0%
		% within CATEG	6,0%	5,4%	,0%	5,3%
	Uma vez por mes	Count	45	2	0	47
		% within N_UFSC	95,7%	4,3%	,0%	100,0%
		% within CATEG	12,3%	3,6%	,0%	10,0%
	Nunca acessei	Count	22	0	0	22
		% within N_UFSC	100,0%	,0%	,0%	100,0%
		% within CATEG	6,0%	,0%	,0%	4,7%
	Outra	Count	6	1	0	7
		% within N_UFSC	85,7%	14,3%	,0%	100,0%
		% within CATEG	1,6%	1,8%	,0%	1,5%
Total		Count	365	56	47	468
		% within N_UFSC	78,0%	12,0%	10,0%	100,0%
		% within CATEG	100,0%	100,0%	100,0%	100,0%

Aqui vale ressaltar que os S.T.A. com 95,7% declararam entrar no *site* www.ufsc.br diversas vezes ao dia (72,3%) ou quase todo dia (23,4%), o que evidencia a utilização no trabalho desenvolvido pelos mesmos em suas atividades diárias. Já os professores para atingir este valor, ou melhor, 94,7% teve-se que considerar aqueles que acessam até de 15 em 15 dias somente.

6.3.2 Quanto à avaliação das dimensões

Tabela 8 - Estatísticas das avaliações das dimensões

Dimensão da Qualidade	N	Minimum	Maximum	Mean	Std. Deviation
AD_01 - Precisão	497	1,7	10,0	7,37	1,490
AD_13 - Confiabilidade	493	,0	10,0	6,99	1,810
AD_20 - Credibilidade	488	1,7	10,0	6,97	1,650
AD_12 - Usabilidade	497	,0	10,0	6,96	1,858
AD_11 - Concisão	492	,0	10,0	6,90	1,800
AD_05 - Segurança	492	,0	10,0	6,86	1,946
AD_06 - Utilidade	496	,0	10,0	6,83	1,746
AD_15 - Valor agregado	492	,0	10,0	6,79	1,739
AD_19 Compreensibilidade	493	,0	10,0	6,79	1,767
AD_03 - Consistência	497	,0	10,0	6,77	1,827
AD_16 - Relevância	490	,0	10,0	6,74	1,748
AD_07 - Atualidade	497	,0	10,0	6,71	2,171
AD_10 - Objetividade	493	,0	10,0	6,40	1,927
AD_14 - Quantidade de dados	495	,0	10,0	6,40	1,900
AD_08 - Disponibilidade	485	,0	10,0	6,29	1,953
AD_17 - Reputabilidade	486	,0	10,0	6,25	1,843
AD_18 - Acessibilidade	491	,0	10,0	6,20	1,979
AD_02 - Navegabilidade	498	,0	10,0	6,19	1,993
AD_04 - Eficiência	498	,0	10,0	6,16	2,074
AD_09 - Completude	494	,0	10,0	6,01	1,794
Valid N (listwise)	440				

Como pode ser visto a Dimensão da Qualidade melhor avaliada foi a Dimensão 1 (Média 7,37) – A informação descrita é correta, confiável e livre de erros (Precisão) enquanto que a de pior avaliação foi a Dimensão 9 (Média 6,01) – A informação está presente; suficientemente ampla e profunda para as tarefas que se tem que realizar (Completude). A dimensão

de maior variação dada pelo desvio padrão (2,171) foi a Dimensão 7 – O *site* está sempre atualizado.

6.3.3 Quanto à importância dada às dimensões

Tabela 9 - Estatísticas da importância dada às dimensões

Importância dada à Dimensão	N	Minimum	Maximum	Mean	Std. Deviation
IDD_01 - Precisão	469	3	10	8,63	1,550
IDD_13 - Confiabilidade	468	2	10	8,56	1,727
IDD_20 - Credibilidade	461	3	10	8,52	1,660
IDD_07 - Atualidade	467	1	10	8,48	1,856
IDD_19 - Compreensibilidade	465	2	10	8,37	1,683
IDD_12 - Usabilidade	469	1	10	8,32	1,746
IDD_02 - Navegabilidade	472	1	10	8,24	1,870
IDD_04 - Eficiência	470	1	10	8,23	1,862
IDD_16 - Relevância	461	2	10	8,13	1,648
IDD_06 - Utilidade	467	1	10	8,12	1,662
IDD_10 - Objetividade	464	1	10	8,11	1,813
IDD_05 - Segurança	460	1	10	8,08	1,928
IDD_15 - Valor agregado	462	2	10	8,05	1,625
IDD_18 - Acessibilidade	466	1	10	7,97	1,818
IDD_14 - Quantidade de dados	468	1	10	7,94	1,749
IDD_09 - Completude	466	2	10	7,88	1,839
IDD_17 - Reputabilidade	460	2	10	7,87	1,795
IDD_11 - Concisão	462	1	10	7,85	1,718
IDD_08 - Disponibilidade	460	1	10	7,84	1,847
IDD_03 - Concistência	468	2	10	7,83	1,668
Valid N (listwise)	429				

Como pode ser visto a importância dada a cada dimensão possui maior média (8,63) para a Dimensão 1 – A informação descrita é correta, confiável e livre de erros (Precisão) enquanto que a de pior avaliação foi a Dimensão 3 (Média 7,83) – A informação é apresentada no mesmo formato e é compatível quando se navega pelo *site* (Consistência). A dimensão de maior variação dada pelo desvio padrão (1,928) foi a Dimensão 5 – O acesso à informação é restrito de forma adequada para manter a segurança.

6.3.4 Quanto à percepção da qualidade

Tabela 10 - Avaliação da percepção da qualidade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Boa	168	33,5	36,7	36,7
	Ruim	290	57,8	63,3	100,0
	Total	458	91,2	100,0	
Missing	9	44	8,8		
Total		502	100,0		

A tabela apresenta na percepção dos usuários a Qualidade do *site* www.ufsc.br. Baseado no critério estabelecido 57,8% dos usuários declarou sua insatisfação com a Qualidade da Informação apresentada no *Site* enquanto que 33,8% afirmaram ver qualidade no mesmo.

6.4 APLICAÇÃO DOS MODELOS

O estudo apresenta três modelos para identificar os parâmetros da qualidade percebida pelos clientes do *site*, quais sejam:

6.4.1 Mod1 – Considerando a categoria e o sexo do cliente

$$\text{Modelo 1: } \log\left(\frac{QP}{1-QP}\right) = \alpha + \beta_i \text{Categoria}_i, \quad i = 1, 2, 3 \quad \text{Equação 17}$$

onde:

QP: Qualidade do *site* percebida pelo usuário (0=Baixa; 1=Alta)

Categoria: Categoria funcional (1=Alunos; 2=Servidores Professores; 3=Servidores Técnico-Administrativos).

Objetivo: Avaliar para a população amostrada a percepção da qualidade do *site* www.ufsc.br para cada uma das categorias e sexos que compõem a comunidade universitária. O interesse é o de inferir se há diferença significativa de percepção da qualidade entre os integrantes da comunidade universitária com base nestas duas características.

Resultados: Nenhum coeficiente se mostrou significativo para explicar a percepção da qualidade. Os resultados são apresentados nas Tabelas 11 e 12.

Tabela 11 - Testes globais para os coeficientes no modelo

		Chi-square	df	Sig.
Step 1	Step	2,986	3	,394
	Block	2,986	3	,394
	Model	2,986	3	,394

Tabela 12 - Estimativas dos coeficientes

		B	S.E.	Wald	df	Sig.	Exp(B)
Step a 1	SEXO(1)	-,125	,235	,284	1	,594	,882
	CATEG			2,604	2	,272	
	CATEG(1)	,019	,390	,002	1	,960	1,020
	CATEG(2)	-,698	,567	1,514	1	,219	,498
	Constant	-,653	,399	2,684	1	,101	,520

a. Variable(s) entered on step 1: SEXO, CATEG.

6.4.2 Mod2 - Considerando os grupos de dimensões

$$\text{Modelo 2: } \log\left(\frac{QP}{1-QP}\right) = \alpha + \beta_1 \bar{C} + \beta_2 \bar{R} + \beta_3 \bar{I} + \beta_4 \bar{A} \quad \text{Equação 18}$$

onde:

QP: Qualidade do *site* percebida pelo usuário (0=Baixa; 1=Alta)

$\bar{C}, \bar{R}, \bar{I} e \bar{A}$: Média das notas atribuídas para cada uma das dimensões avaliadas para as categorias (C=Contextual, R=Representacional, I=Intrínseca e A=Acessibilidade, modelo CRIA).

Objetivo: Avaliar para a população amostrada os efeitos de cada uma das categorias que reúnem as dimensões de qualidade da informação para o *site* www.ufsc.br. O interesse é o de inferir o grupo de dimensões mais relevante para a população.

Resultados: O conjunto de dados foi preparado de forma a não incluir nenhum caso com valores ausentes. Deste modo foram processadas informações de 458 casos coletados. A Tabela 13 apresenta o sumário do modelo processado, em termos da função *loglikelihood*, e os valores dos pseudo R's (Cox & Snell; Nagelkerke). O processo de estimação foi concluído na iteração de número 6 dado que as trocas nas estimativas dos parâmetros atingiram o valor limite (0,001).

Tabela 13 - Sumário do modelo

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	360,005	,410	,561

Em termos de classificação, o modelo atingiu um percentual global de 83,2%, conforme é apresentado na Tabela 14. Os coeficientes encontrados estão descritos na Tabela 15.

Tabela 14 - Classificação da variável dependente^a

			Predicted		
			Perc_Qualidade		Percentage Correct
			Ruim	Boa	
Step 1	Perc_Qualidade	Ruim	260	30	89,7
		Boa	47	121	72,0
Overall Percentage					83,2

a. Valor do corte igual a 0,5

Tabela 15 - Estimativas dos coeficientes

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Contextual	,528	,173	9,280	1	,002	1,696
	Representacional	,378	,165	5,277	1	,022	1,460
	Intrínseca	,439	,195	5,086	1	,024	1,552
	Acessibilidade	,652	,135	23,500	1	,000	1,920
	Constante	-14,151	1,345	110,628	1	,000	,000

a. Variable(s) entered on step 1: CO, RE, IN, AC.

Todos os coeficientes encontrados são significativos a 5%. A categoria de dimensões de qualidade que mais contribui para que aumente a percepção da qualidade do *site* é a Acessibilidade. A que menos importa é a Intrínseca, própria da informação, ou seja, a forma como a informação está sendo buscada ou gerada.

6.4.3 Mod3 - Considerando os grupos de dimensões e as categorias

Mod3A – Considerando a categoria Alunos

O conjunto de dados foi preparado com informações obtidas de 354 estudantes, de forma a não incluir nenhum caso com valores ausentes. As estatísticas função $-2 \times \text{loglikelihood}$, R^2 de Cox e Snell; e R^2 de Nagelkerke obtidas foram respectivamente 293,958; 0,401 e 0,544. O percentual de acerto global obtido pelo modelo foi de 82,4%.

A Tabela 16 mostra os valores e estatísticas obtidos para os coeficientes do modelo sem efeitos mistos. Todas as variáveis preditoras são estatisticamente significativas a 5%.

Tabela 16 - Estimativas dos coeficientes do modelo 2A - alunos

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Contextual	,482	,189	6,539	1	,011	1,620
	Representacional	,349	,173	4,072	1	,044	1,417
	Intrínseca	,495	,213	5,405	1	,020	1,641
	Acessibilidade	,638	,150	18,008	1	,000	1,893
	Constante	-13,903	1,486	87,552	1	,000	,000

a. Variable(s) entered on step 1: CO, RE, IN, AC.

A matriz de correlação dos coeficientes é apresentada na Tabela 17. Nota-se que todas as correlações obtidas são negativas. Uma das causas está relacionada à codificação adotada neste trabalho (valores maiores correspondem a uma avaliação mais negativa da característica).

Os sinais negativos, portanto, implicam o fato de que todas as categorias estão na mesma direção. Não há conflito para a categoria dos estudantes entre as dimensões que avaliam a qualidade do *site* www.ufsc.br.

Tabela 17 - Matriz de correlações dos coeficientes do modelo 2A

		Constante	Contextual	Represent.	Intrínseca	Acessibilidade
Step 1	Constante	1,000	-,247	-,319	-,222	-,424
	Contextual	-,247	1,000	-,084	-,464	-,119
	Representacional	-,319	-,084	1,000	-,384	-,076
	Intrínseca	-,222	-,464	-,384	1,000	-,093
	Acessibilidade	-,424	-,119	-,076	-,093	1,000

Mod3B – Considerando a categoria Professores

O conjunto de dados foi preparado com informações obtidas de 58 professores, de forma a não incluir nenhum caso com valores ausentes. As estatísticas função $(-2) \times \log\text{likelihood}$, R^2 de Cox e Snell; e R^2 de Nagelkerke obtidas foram respectivamente 14,125; 0,595 e 0,869. O percentual de acerto global obtido pelo modelo foi de 91,2%.

A Tabela 18 mostra os valores e estatísticas obtidos para os coeficientes do modelo.

Tabela 18 - Estimativas dos coeficientes do modelo 2B - Servidores professores

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Contextual	4,400	1,957	5,054	1	,025	81,470
	Representacional	1,555	1,582	,967	1	,326	4,735
	Intrínseca	-1,833	1,281	2,048	1	,152	,160
	Acessibilidade	,525	,748	,492	1	,483	1,690
	Constante	-32,430	12,669	6,552	1	,010	,000

a. Variable(s) entered on step 1: CO, RE, IN, AC.

A única variável preditora estatisticamente significativa foi o grupo Contextual. Esta variável é responsável pela percepção de qualidade avaliada pelos professores. É interessante ressaltar que o grupo de dimensões que mede a dimensões intrínsecas da qualidade (grupo Intrínseca) apresenta um coeficiente negativo. Neste grupo estão dimensões como precisão, confiabilidade, disponibilidade, credibilidade, reputabilidade e objetividade. Assim, a categoria dos professores debita na ausência destas dimensões a não-qualidade do *site*.

A matriz de correlação dos coeficientes é apresentada na Tabela 19. Nota-se a ausência de correlação entre os coeficientes dos grupos de dimensões associadas à Acessibilidade e ao grupo Contextual. Destaca-se, ainda, para a categoria dos professores, a presença de alternância de sinais de correlação, ou seja, existem grupos de dimensões de qualidade que não atendem aos critérios subjetivos de qualidade utilizados pelos professores.

Tabela 19 - Matriz de correlações dos coeficientes do modelo 2B

		Constant	Contextual	Represent.	Intrínseca	Acessibilidade
Step 1	Constante	1,000	-,747	-,513	,490	-,253
	Contextual	-,747	1,000	,216	-,737	,006
	Representacional	-,513	,216	1,000	-,586	-,399
	Intrínseca	,490	-,737	-,586	1,000	,263
	Acessibilidade	-,253	,006	-,399	,263	1,000

Mod3C – Considerando a categoria Técnico-Administrativos

O conjunto de dados foi preparado com informações obtidas de 46 servidores técnico-administrativos, de forma a não incluir nenhum caso com valores ausentes. As estatísticas função $(-2) \times \loglikelihood$, R^2 de Cox e Snell; e R^2 de Nagelkerke obtidas foram respectivamente 39,379; 0,315 e 0,433. O percentual de acerto global obtido pelo modelo foi de 76,7%.

A Tabela 20 mostra os valores e estatísticas obtidas para os coeficientes do modelo. O modelo não apresentou variáveis preditoras significativas.

Tabela 20 - Estimativas dos coeficientes do modelo 2C - Servidores técnico-administrativos

		B	S.E.	Wald	df	Sig.	Exp(B)
Step a 1	Contextual	-,181	,562	,104	1	,747	,834
	Representacional	,837	,702	1,424	1	,233	2,311
	Intrínseca	,222	,759	,085	1	,770	1,248
	Acessibilidade	,634	,484	1,719	1	,190	1,885
	Constante	-10,869	3,763	8,345	1	,004	,000

a. Variable(s) entered on step 1: CO, RE, IN, AC.

Não houve nenhuma variável estatisticamente significativa para os servidores técnico-administrativos. Aqui, o grupo de dimensões que mede a dimensão contextual da qualidade apresenta um coeficiente negativo. Neste grupo estão dimensões como relevância, atualidade, completude, quantidade de informação e valor adicionado. Assim, a categoria dos servidores técnico-administrativos indica a ausência destas dimensões quando avalia a qualidade do *site*.

A matriz de correlação dos coeficientes é apresentada na Tabela 21. Nota-se a ausência de correlação entre os coeficientes dos grupos de dimensões associadas à Acessibilidade e ao grupo Contextual. Destaca-se, ainda, para a categoria dos servidores técnico-administrativos, a presença de alternância de sinais de correlação, ou seja, existem grupos de dimensões de

qualidade que não atendem aos critérios subjetivos de qualidade utilizados pelos servidores.

A Tabela 21 apresenta a matriz de correlações para os coeficientes obtidos.

Tabela 21 - Matrix de correlações dos coeficientes do modelo 2C

		Constante	Contextual	Represent.	Intrínseca	Acessibilidade
Step 1	Constante	1,000	-,141	-,422	,032	-,395
	Contextual	-,141	1,000	-,335	-,313	-,024
	Representacional	-,422	-,335	1,000	-,500	,183
	Intrínseca	,032	-,313	-,500	1,000	-,501
	Acessibilidade	-,395	-,024	,183	-,501	1,000

6.5 MODELO DE OTIMIZAÇÃO DIFUSA

Considerem-se os coeficientes de regressão apresentados na Tabela 18, p. 76, que é o modelo que melhor se ajustou só com as dimensões. A equação abaixo é obtida pelo uso dos coeficientes e será utilizada no modelo de otimização difusa.

$$Y = -14,151 + 0,528C + 0,378R + 0,439I + 0,652A \quad \text{Equação 19}$$

A análise dos coeficientes do modelo logit indica que a variável que mais contribui para aumentar a probabilidade de uma avaliação positiva do *site* é a avaliação atribuída ao conjunto de dimensões que definem a acessibilidade.

Entretanto, estes coeficientes são imprecisos no sentido de mensurar o efeito que acarretam nos processos cognitivos dos usuários. Então, pode-se admitir que o valor de probabilidade obtido também represente uma aproximação vaga em torno do real valor. Assim, tem-se que também os valores dos coeficientes são imprecisos e vagos.

Considerando que se deseja obter um nível de aspiração de 90% na representatividade destes coeficientes, bem como da probabilidade de classificação correta a respeito da qualidade do *site*, e que a tolerância máxima admitida é de 70%, o problema pode ser solucionado como um problema de minimização de função e de coeficientes difusos, sujeitos a estas duas restrições, conforme o modelo apresentado na Função 8 do capítulo 3, p.45.

Além destes aspectos, admita-se que para cada ponto de valoração das notas que medem as diversas categorias de dimensões da qualidade existam valores monetários associados. Por exemplo, admita-se que cada ponto de avaliação custe 25 unidades monetárias para as dimensões associadas à categoria **Intrínseca**; 20 para a categoria **Representacional**; 25 para a **Contextual** e 30 para a de **Acessibilidade**. Considere-se, ainda, que estão disponíveis 350 unidades no total.

Com relação às notas obtidas, se deseja manter o valor mínimo de 8 para a categoria de Acessibilidade, uma vez que esta categoria é a que mais contribui para aumento na probabilidade de uma boa avaliação da qualidade. Evidentemente todas as notas devem estar limitadas a 10.

Sob estas condições, tem-se o seguinte problema de minimização difusa (que engloba os princípios admitidos para os modelos de otimização difusa propostos no capítulo 5):

$Min \theta$

sujeito a :

$$0,7\theta + 0,528C + 0,378R + 0,439I + 0,652A \geq 15,051 \quad \text{Equação 20}$$

$$-245\theta + 25C + 20R + 15I + 30A \leq 315$$

$$0 \leq C, R, I \leq 10$$

$$8 \leq A \leq 10$$

que resulta nos seguintes valores que devem ser alcançados pelas categorias de dimensões:

Contextual = 6,26

Representacional = 0,0

Intrínseca = 10

Acessibilidade = 10

com um nível de tolerância de 0,7 e de aspiração de 0,9 e as funções de pertinência dadas pela Função 7 do capítulo 3, p.45.

6.6 DISCUSSÃO DOS RESULTADOS

Os resultados deste trabalho demonstram que a percepção da qualidade tem diferenças, estatisticamente significativas, pelos grupos de dimensões estabelecidos considerando os usuários já caracterizados. Assim é possível através de programação difusa obter para um nível de aspiração desejado na representatividade dos coeficientes, bem como da probabilidade de classificação correta a respeito da qualidade do *site*, considerando uma tolerância máxima admitida, a solução do problema pela minimização de função e considerando coeficientes difusos, sujeitos às restrições.

7 CONCLUSÕES E RECOMENDAÇÕES

7.1 CONCLUSÕES

O trabalho apresenta um modelo para avaliar a percepção da qualidade de um site tendo como base características dos usuários e dimensões de qualidade.

O modelo proposto identifica inicialmente as características da comunidade de usuários de um determinado site. A identificação destas características é feita com o auxílio de um estudo descritivo exploratório de dados. Esta análise é importante porque descreve a comunidade e permite identificar perspectivas cognitivas e sociológicas que possam vir a interferir na percepção da qualidade.

Identificado o usuário ou a comunidade, o modelo considera o estabelecimento de dimensões de qualidade da informação que devem estar presentes no site.

Uma vez estabelecida a comunidade a ser analisada bem como o site que atende a mesma, um instrumento de coleta de dados é desenvolvido de forma a permitir a caracterização dos membros desta comunidade além de levantar as valorações atribuídas às diferentes dimensões consideradas.

Como retorno tem-se um perfil da comunidade e das dimensões da qualidade que esta considera relevantes para o seu site.

Através de um procedimento pré-estabelecido estas dimensões são estratificadas de acordo com a relevância dada pela comunidade.

Para identificar as dimensões da qualidade que devem ser consideradas o mecanismo utilizado para recuperar a informação sobre a qualidade do site mede o sentimento dos usuários quanto a realidade atual de cada dimensão; a expectativa idealizada e a percepção global da qualidade.

Para identificar como características da comunidade influenciam na percepção da qualidade o modelo proposto neste trabalho utiliza a técnica de regressão logística. De igual modo, esta técnica é adotada para identificar as dimensões relevantes.

Os resultados obtidos pelos modelos de regressão logit propostos constituem-se em orientações que podem estar limitadas quer por questões técnicas quer por razões de limitação de recursos.

Assim, o modelo prevê, então a utilização de um procedimento de otimização linear difusa, no qual, as restrições podem ser os resultados obtidos pelos modelos de regressão ou considerar que os efeitos dos regressores são melhor representados por números difusos. Deste modo a modelo proposto no trabalho apresenta duas abordagens para o tratamento deste problema de decisão.

Para avaliar o modelo proposto foi desenvolvida uma aplicação junto a comunidade universitária da UFSC. Esta comunidade, dividida em três segmentos – servidores docentes; servidores técnico administrativos e corpo discente, foi consultada a respeito da qualidade do portal www.ufsc.br.

Foram avaliadas 20 dimensões de qualidade. Estas de acordo com o modelo proposto foram então agrupadas em quatro categorias: Acessibilidade à Informação; Representação da Informação; Contextualidade da Informação e Qualidade Intrínseca.

Entre as características individuais não foram encontradas diferenças significativas para percepção da qualidade. Entretanto, quando considerada a comunidade como um todo, as características de qualidade do site mostraram-se significativas.

Pode-se dizer, neste caso que o grupo de dimensões mais relevante foi aquele que reunia as dimensões associadas ao grupo Acessibilidade. De fato, em uma instituição universitária o que se busca é informação, portanto, este grupo de características de qualidade é o mais representativo da qualidade (ou não qualidade) do site.

É apresentado, ainda, para o estudo piloto um pequeno protótipo para níveis de aspiração e de tolerância máxima a serem obtidos quando o processo está sujeito a limitações. Na aplicação realizada foram consideradas restrições monetárias e níveis mínimos de avaliação da qualidade.

Como resultado, o modelo aponta para priorizar as dimensões de qualidade associadas às características de Acessibilidade; Qualidade Intrínseca, Contextual, ignorando as dimensões de qualidade da informação associadas à Representacional.

7.2 RECOMENDAÇÕES PARA FUTUROS TRABALHOS

O modelo proposto está centrado na percepção da qualidade que o usuário apresenta e na expectativa idealizada.

Para trabalhos futuros, sugere-se incorporar o aspecto temporal para mensurar a evolução desta percepção ao longo do tempo.

As técnicas utilizadas para construir um site têm impacto direto na qualidade do mesmo. Neste trabalho estas influências não são consideradas. Propõe-se incorporar ao presente modelo alternativas, que considerem diferentes técnicas de construção, procurando identificar se existem mecanismos mais adequados para a construção de sites que ampliem a percepção da qualidade.

Uma característica relevante da comunicação medida pela tecnologia é a interatividade, assim sugere-se para trabalhos futuros considerar mecanismos que ao interagirem com o usuário permitam identificar de forma instantânea a qualidade do site. Tal pode ser obtido com o auxílio de agentes de software especificamente desenvolvidos para tal fim.

Finalmente, sugere-se a aplicação do modelo proposto neste trabalho a outras comunidades. Por exemplo, ao aplicar-se a outras instituições acadêmicas de ensino, o modelo apresentado poder-se-ia estudar também o efeito da instituição na percepção global da qualidade, esta informação poderia ser útil para direcionar políticas globais de desenvolvimento das instituições.

8. REFERÊNCIAS

- ALEXANDER, J.E.; TATE, M.A., **Web wisdom: How to evaluate and create information quality on the Web**. Mahwah, NJ: Lawrence Erlbaum, 1999.
- ALLAN, J., CARTERETTE, B., LEWIS J., **When Will Information Retrieval Be “Good Enough”? User Effectiveness As a Function of Retrieval Accuracy**, SIGIR’05, August 15–19, 2005, Salvador, Brazil.
- BACHARACH, M., Board, O., **The quality of information in electronic groups**, Netnomics 4: 73–97, 2002. Kluwer Academic Publishers. Printed in the Netherlands.
- BELMANN, R.E., ZADEH, L.A., **Decision Making in a Fuzzy Environment**. Management Science, Vol. 17: 141-164. 1970.
- BOUZEGHOUB, M., PERALTA, V., **A Framework for Analysis of Data Freshness**, IQIS 2004 Maison de la Chimie, Paris, France.
- BRAUMANDL, R., KEMPER, A., KOSSMANN, D., **Quality of Service in an Information Economy**, University of Passau, Germany and ACM Transactions on Internet Technology, V. 3, No. 4, November 2003, Pages 291–333.
- BUGAJSKI, J., GROSSMAN, R.L., SUMNER, E., TANG, Z., **An Event Based Framework for Improving Information Quality That Integrates Baseline Models, Causal Models and Formal Reference Models**, Bearing Point IQIS 2005, June 17, 2005, Baltimore, MD, USA.
- CAPPIELLO C.,FRANCALANCI, C.,PERNICI, B., **Data quality assessment from the user’s perspective**, IQIS 2004 Maison de la Chimie, Paris, France.
- CAPPIELLO, C., FRANCALANCI, C., PERNICI, B., **“Time-Related Factors of Data Quality in Multichannel Information Systems”**, JMIS. V. 20 No. 3 , Winter 2004 , p. 71-91
- DAVIDSON, I., GROVER, A., SATYANARAYANA, A., TAYI, G.K., **A General Approach to Incorporate Data Quality Matrices into Data Mining Algorithms**, KDD’04, August 22–25, 2004, Seattle, Washington, USA.

De Giacomo, G., LEMBO, D., LENZERINI, M., ROSATI, R., **Tackling Inconsistencies in Data Integration through Source Preferences**, IQIS 2004 Maison de la Chimie, Paris, France

DEDEKE, A., **A conceptual framework for developing quality measures for information systems**. Proceedings of 5th International Conference on Information Quality. p. 126–128, 2000.

DeLONE, W.D., McLEAN, E.R. **"The DeLone and McLean Model of Information Systems Success: A Ten-Year Update"**, Journal of Management Information Systems, V. 19 No. 4, Spring 2003, p. 9-30

DEVANSHU, D., WEE, K. NG, SOURAV S. B., **A Survey of Web Metrics**, Nanyang Technological University, ACM Computing Surveys, V. 34, No. 4, December 2002, p. 469–503.

EPPLER, M., MUENZENMAYER, P., **Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology**. In: Fisher, C., Davidson, B.N. (Ed.). Proceedings of the 7th International Conference on Information Quality, MIT, 2002, p. 187-196, 2002.
<http://www.knowledgemedia.org/modules/pub/view.php/knowledgemedia-27>. Acesso em: 12 jun 2006.

GARRITY, E. J., SANDERS, G.L., **Information System Success Measurement**, (Ed). Idea Group Publishing, p.1-12, 1998.
<http://portal.acm.org/citation.cfm?id=275785&dl=ACM&coll=&CFID=15151515&CFTOKEN=6184618>. Acesso em: 19 jun 2006.

GERTZ, M., OZSU, M. T., SAAKE G., SATTler, Kai-Uwe, **Report on the Dagstuhl Seminar "Data Quality on the Web"**, SIGMOD Record, V. 33, No. 1, March 2004

GONÇALVES, C.F.F., **Ergonomia e Qualidade do Serviço Bancário: Uma Metodologia de Avaliação**. Doutorado no PPGEp. UFSC, 1995.

GRAY, M., **Growth and Usage of the Web and the Internet**,
<http://www.mit.edu/people/mkgray/net/web-growth-summary.html> Acesso em: 13 jun 2006.

GUDWIN, R.R., GOMIDE, F.A.C., PEDRYCZ W., **Context Adaptation in Fuzzy Processing and Genetic Algorithms** – International Journal of Intelligent Systems vol. 13, n. 10/11, pp. 929-948, Oct./Nov. 1998.

HARDY, H., **The History of the Net**. Master's Thesis, School of Communications, Grand Valley State University.
<http://www.ocean.ic.net/ftp/doc/nethist.html> Acesso em: 28 mai 2006.

HARDY, I., **The Evolution of ARPANET email**. History Thesis, UC Berkeley.
http://server.berkeley.edu/virtual-berkeley/email_history Acesso em: 15 jun 2006.

_____. Como medir o resultado de um site?
<http://commento.cartello.com.br/?m=200506>. Acesso em: 22 jun 2006.

_____. Linha do tempo da Internet.
<http://homepages.dcc.ufmg.br/~mlbc/cursos/internet/historia/TimeLine.html>
 . Acesso em 23 jun 2006.

IIVARI, J., **An Empirical Test of the DeLone-McLean Model of Information System Success**, The DATA BASE for Advances in Information Systems - Spring 2005 (V. 36, No. 2)

ISO/IEC 9126 - Tecnologia de informação – Avaliação de produto de software – Características de qualidade e diretrizes para o seu uso, Brasil. Versão brasileira da norma ISO/IEC 9126. (**NBR 13596**)

JIANG, J.J., KLEIN G., TESCH, D., CHEN H.G., **A method for measuring service quality that includes both the user and IS service provider perspectives**. Closing the User and Provider Service QUALITY GAP, February 2003/V. 46, No. 2 Communications of the ACM.

KAHN, B. K., STRONG, D. M., WANG, R. Y., **Information Quality Benchmarks: Product and Service Performance**, Communications of the ACM, V. 45, No. 4 84-192, 2002.

KATERATTANAKUL, P.; SIAU, K., **Measuring information quality of web sites: Development of an instrument**. Proceedings of the 20th international conference on Information Systems. Charlotte, North Carolina; p.279–285, 1999.

KELLY, B., VIDGEN, R., **A Quality Framework for Web Site Quality: User Satisfaction and Quality Assurance**. WWW 2005, May 10-14, 2005 Chiba, Japan.

KLEIN, B.D., **User Perceptions of Data Quality: Internet and Traditional Text Sources**, Journal of Computer Information Systems, (41:4), 2001, p. 5-15.

KLEIN, B.D., **When do users detect information quality problems on the World Wide Web?** Eighth Americas Conference on Information Systems, 2002, p. 1101-1103.

KLIR, G.J. FOLGER, T.A., **Fuzzy Sets, Uncertainty, and Information.** Englewood Cliffs: Prentice Hall 1988.

KNIGHT, S. A., BURN, J., **Developing a Framework for Assessing Information Quality on the World Wide Web.** Informing Science Journal. v. 8, 2005.

LAU, T.; HORVITZ, E., **Patterns of Search: Analyzing and Modeling Web Query Refinement.** Proceedings of User Modeling 99. Banff, Canada, p. 119-128, 1999.

LEE, Y.W. "Crafting Rules: Context-Reflective Data Quality Problem Solving", Journal of Management Information Systems, V. 20 No. 3, Winter 2004, p. 93-119

LIMA, R. et al., **Evaluating web sites for an educational environment target for cardiology,** The 3rd European Software Measurement Conference. Espanha, 2000.

LUEBBERS, D., GRIMMER, U., JARKE, M., **Systematic Development of Data Mining-Based Data Quality Tools;** Proceedings of the 29th VLDB Conference, Berlin, Germany, 2003

MADNICK, S. E., WANG, R. Y., XIAN, X., "The Design and Implementation of a Corporate Householding Knowledge Processor to Improve Data Quality", JMIS, V. 20 No. 3 , Winter 2004 , p. 41 - 69

MAJKI 'C, Zoran, **A General Framework for Query Answering in Data Quality-based Cooperative Information Systems,** IQIS 2004 Maison de la Chimie, Paris, France

MISSIER, P., EMBURY, S., **Provider issues in quality constrained data provisioning,** School of Computer Science The University of Manchester, UK, IQIS 2005, June 17, 2005, Baltimore, MD, USA.

MISSIER, P., LALK, G., VERYKIOS, V., GRILLO, F., LORUSSO T., ANGELETTI P., **Improving Data Quality in Practice: A Case Study in the Italian Public Administration,** Distributed and Parallel Databases, Springer Netherlands V. 13, N. 2 , March 2003, p. 135-160

MOCKUS, A., ZHANG, P., LI, P.L., **Predictor of Customer Perceived Software Quality,** ICSE '2005, May 15-21 2005, St. Louis Missouri, USA.

NAUMANN, F., ROLKER, C., **Assessment methods for information quality criteria**, Proceedings of 5th International Conference on Information Quality. p. 148–162, 2000.

NBR ISO 8402/1994 - Gestão da qualidade e garantia da qualidade, <http://www.indg.com.br/info/glossario/glossario.asp?n>. Acesso em: 23 jun 2006.

NEHMY, R. M. Q., PAIM, I., **A desconstrução do conceito de "qualidade da informação"**. <http://www.ibict.br/cienciadainformacao/include/getdoc.php?id=701&article=385&mode=pdf>. Acesso em: 19 abr 2006.

OLSON, J. E., **"Data Quality – The Accuracy Dimension"**, Morgan Kaufmann Publishers, 2003.

PACHECO, R. C. S., **A Hybrid Intelligent System for Diagnosing and Solving Financial Problems**. Doutorado PPGEF, UFSC, 1996.

PASTERNAK, B.A.; VISCIO, A. J., **Mining the riches in knowledge**. The centerless corporation: A new model for transforming your organization for growth and prosperity. New York: Simon & Schuster. Chapter 4. p. 93-127, 1998.

PIERCE, E. M., **Assessing Data Quality With Control Matrices**, February 2004/V. 47, No. 2 Communications of the ACM.

PIPINO, L. L., LEE, Y. W., Wang, R. Y., **Data Quality Assessment**, Communications of the ACM, April 2002 V. 45, No. 4.

PON, R. K., CÁRDENAS, A. F., **Data Quality Inference**, IQIS 2005, June 17, 2005, Baltimore, MD, USA.

PRESSMAN, R. S., **Software Engineering: A Practitioner's Approach**. McGraw Hill, 2001.

REIS, T.P.C.; CASTRO, J. F. B.; OLSINA, L. A., **Medição de Qualidade de Aplicações Web na Fase de Requisitos**. In: Simpósio Brasileiro de Engenharia de Software. Gramado, RS, 2002, p. 162-174.

RIVERA, T., TATE, A., WILL, S. A., **Actively Involving Our Information Development Teams with Clients**, SIGDOC'03, October 12–15, 2003, San Francisco, California, USA.

RYAN, N., TODD, P.A., WIXON, G.B.H., **"Antecedents of Information and System Quality: An Empirical Examination Within the Context of Data Warehousing"**, Journal of Management Information Systems, and System Quality, V. 21 No. 4 , Spring 2005 , p. 199–236

SEIN, U.T.; WAHEED, A., **National Health Accounts: Policy Brief on Concepts and Approaches**, WHOSEA Regional Health Forum, v. 7, n. 2, 2003. <http://w3.whosea.org/rhf/rhf7-2/HealthPolicy.htm>. Acesso em: 22 Jun 2006.

SHANKS, G.; CORBITT, B., **Understanding data quality: Social and cultural aspects**, Proceedings of the 10th Australasian Conference on Information Systems. p. 785, 1999.

SMART, K. L., **Assessing Quality Documents**, ACM Journal of Computer Documentation August 2002/V. 26, No. 3

STAMEY, J.W. Jr., HONEYCUTT, T. L., BLANCHARD, S., **Edward Tufte Meets Christopher Alexander**, SIGDOC'05, September 21–23, 2005, Coventry, United Kingdom.

TAYI, G.K.; BALLOU, D.P., **Examining Data Quality**, Communications of ACM, v. 41, n. 2, p. 54-57, 1998.

TROCHIM, W.M.K. **Evaluating Sites Web**, 1999, <http://trochim.human.cornell.edu/Webeval/Webintro/Webintro.htm>. Acesso em: 12 jun 2006.

WANG, R.Y.; LEE, Y.; PIPINO, L.; STRONG, D., **Manage Your Information as a Product**, Sloan Management Review. Summer 1998, v. 39, n. 4. p. 95-105.

WANG, R.Y., ZIAD, M., LEE. Y.W., **"Data Quality"**, Springer Verlag, 2002

WANG, R.Y.; STRONG, D.M. Beyond Accuracy: **What Data Quality Means to Data Consumers**, Journal of Management Information Systems (12:4), 1996, p. 5-34.

WEISS, E.H., **The Metaphysics of Information Quality: Comments on Producing Quality Technical Information**, ACM Journal of Computer Documentation August 2002/V. 26, No. 3

ZADEH, L. A., **Fuzzy sets**, Information and Control. v.8, p. 338-353, 1965.

ZEIST, R.H.J.; HENDRIKS, P.R.H., **Specifying software quality with the extended ISO model**, Software Quality Management IV – Improving Quality, BCS, 145-160, 1996.

ZEMANKOVA, M., KANDEL, A, **Fuzzy Relational Database: A Key to Expert Systems**, Verlag TNV Rheinland, Definition of additive possibilistic expected value; hard to find in text? 1984.

ZHANG, Dazhi; OUYANG, He; and LEE, E et al., **On Fuzzy Random Sets and Their Mathematical Expectations**, Information Sciences, v. 72, p.123-142, 1993.

ZHOU, Y., CROFT, W. B., **Document Quality Models for Web Ad Hoc Retrieval**, CIKM'05, October 31-November 5, 2005, Bremen, Germany.

ZHU, X., GAUCH, S. (2000). **Incorporating quality metrics in centralized -distributed information retrieval on the World Wide Web**. Proceedings of the Twenty- Third Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 288-295.

ZIMMERMAN, H.J., **Fuzzy Sets, Decision Making and Expert Systems**, Kluwer, B.V. Deventer, The Netherlands, The Netherlands, p. 198-215 1987.

Apêndice



UNIVERSIDADE FEDERAL DE SANTA CATARINA

CENTRO TECNOLÓGICO

Programa de Pós-graduação em Engenharia de Produção

Bom dia. Meu nome é Manuel Rosa de Oliveira Lino. Sou professor do Departamento de Informática e Estatística.

Este instrumento faz parte do trabalho correspondente à tese que estou defendendo no Programa de Pós-graduação em Engenharia de Produção, com o título "Ferramentas para análise de qualidade da informação em *WebSites* através de modelos Difusos".

A investigação é sobre o **Site da Universidade Federal de Santa Catarina**. (<http://www.ufsc.br>)

Gostaria que, por favor, preenchesse o questionário abaixo. As respostas são confidenciais e serão utilizadas, tão somente, para modelagem de ferramentas de análise de *Sites*.

Perfil do Usuário	Centro	Categoria: Alunos → Curso: Professores → Depto: Serv. Téc. Adm. → Depto:
	Idade () anos	
Sexo () 1. F ; 2. M		

<i>Atenção no item Local você deve assinalar, com as alternativas à direita do item, nos três locais possíveis a frequência de acessos, exemplo:</i> <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> Na UFSC: diversas vezes ao dia → assinale Na residência: Uma vez por semana → assinale Em outro local: Uma vez por mês → assinale </div> <div style="width: 45%;"> Na UFSC (1) Na residência (4) Em outro local (6). </div> </div>		<i>Assinale o tipo de acesso na UFSC e na Residência se possuir.</i>				
Local, Frequência e Tipo de acesso ao Site.	<table style="width: 100%;"> <tr> <td style="width: 30%;">Local:</td> <td style="width: 70%;"> 1. Diversas vezes ao dia; 2. Praticamente todo dia; 3. Em torno de três vezes por semana; 4. Uma vez por semana; 5. De 15 em 15 dias; 6. Uma vez por mês; 7. Nunca acessei; 8. Outro. Quantas? </td> </tr> <tr> <td> Na UFSC (); Na residência (); Em outro local (). </td> <td></td> </tr> </table>	Local:	1. Diversas vezes ao dia; 2. Praticamente todo dia; 3. Em torno de três vezes por semana; 4. Uma vez por semana; 5. De 15 em 15 dias; 6. Uma vez por mês; 7. Nunca acessei; 8. Outro. Quantas?	Na UFSC (); Na residência (); Em outro local ().		Tipo: Na UFSC (); 1. Cabo 2. WireLess. Na Residência (); 1. Discada 2. Alta Velocidade.
Local:	1. Diversas vezes ao dia; 2. Praticamente todo dia; 3. Em torno de três vezes por semana; 4. Uma vez por semana; 5. De 15 em 15 dias; 6. Uma vez por mês; 7. Nunca acessei; 8. Outro. Quantas?					
Na UFSC (); Na residência (); Em outro local ().						

Atenção o instrumento prevê uma análise em duas dimensões. Na primeira você deverá atribuir um único conceito desde *Excelente* a *Péssimo* para cada item da tabela abaixo. Na segunda você deverá votar de Zero a Dez que importância dá ao item no acesso ao *Site*.
Por favor, não deixe nenhum item sem uma resposta.

Item	Neste quadro você deve assinalar uma alternativa para cada um dos itens de 1 a 21 além de atribuir uma nota de Zero as Dez na última coluna, para cada um, baseado na importância que você dá ao item. Pode haver várias notas iguais a Dez como, em fim, qualquer outra nota. Sobre o Site – página principal da UFSC – http://www.ufsc.br	Excelente	Muito bom	Bom	Regular	Fraco	Ruim	Péssimo	Valor dado ao item, de Zero a Dez
1	A informação descrita é correta, confiável e livre de erros.								
2	Os dados são facilmente encontrados e existem ligações adequadas.								
3	A informação é apresentada no mesmo formato e é compatível quando se navega pelo <i>Site</i> .								
4	Os dados disponíveis no <i>Site</i> facilitam encontrar rapidamente a informação necessária.								
5	O acesso à informação é restrito de forma adequada para manter a segurança.								
6	A informação é importante e útil para tarefas que se tem que realizar.								
7	O <i>Site</i> está “sempre” atualizado.								
8	A informação é fisicamente acessível (possibilidade de download).								
9	A informação está presente; suficientemente ampla e profunda para tarefas que se tem que realizar.								
10	A informação é imparcial e não distorcida.								
11	A informação está representada compactamente sem ser rebuscada demais.								
12	A informação é clara e de fácil utilização.								
13	A informação é correta e confiável.								
14	A quantidade ou volume de dados disponíveis é apropriado.								
15	A informação oferece vantagem com o seu uso.								
16	A informação é aplicável e útil para tarefas que se tem que realizar.								
17	A informação é altamente classificada em termos de fonte ou de conteúdo.								
18	A informação está disponível ou, fácil e rapidamente recuperável.								
19	Os dados são claros, sem ambigüidade e facilmente compreensíveis.								
20	A informação é classificada como crível e verdadeira.								
21	Que resposta daria se lhe perguntassem sobre a qualidade do <i>Site</i> da UFSC.								xxxxx

Se você tivesse a oportunidade de sugerir alguma mudança no *Site* da UFSC o que você indicaria (use o ESPAÇO ABAIXO).